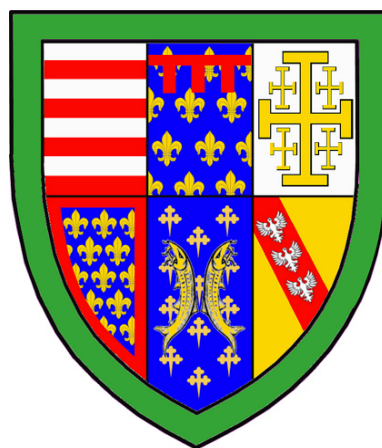


STUDYING THE CLONAL ORIGINS OF DRUG RESISTANCE IN HUMAN BREAST CANCERS



John William Cassidy, MSci

Cancer Research UK Cambridge Institute

Department of Oncology

University of Cambridge

This dissertation is submitted for the degree of *Doctor of Philosophy*

Queens' College

September 2018

STUDYING THE CLONAL ORIGINS OF DRUG RESISTANCE IN HUMAN BREAST CANCERS

John William Cassidy, MSci

Summary

Despite remarkable advances in our understanding of the drivers of human malignancies, new targeted therapies often fail to show sufficient efficacy in clinical trials. Indeed, the cost of bringing a new agent to market has risen substantially in the last several decades, fuelled partly by lack of efficacy in late phase clinical trials. Even in cases where a new agent is deemed 'successful', the development of resistance is often seen as inevitable and clinical responses can be fleeting. Typically, resistance to targeted therapies is thought to arise from pre-existing populations within the tumour, rather than from *de novo* evolution, yet few studies have experimentally tested this understanding. Indeed, recent reports in the literature have described epigenetically regulated drug tolerant populations within cancers, defined by cell-cycle regulation and/or quiescent repopulation dynamics, drug induced chromatin remodelling or differential transcription factor binding, that can be transient or permanent in nature. This thesis will outline experiments using high complexity molecular barcodes to trace the fate of individual cellular clones in the development of drug resistance. With this technique, cellular clones can be uncoupled from their genomic backgrounds, giving a new depth to our understanding of clonal selection in cancer. In particular, high complexity barcodes are used to identify a pre-existing tamoxifen resistant population in the MCF7 cell line. This resistance phenotype is then linked to the induction of embryonic transcription factor OCT4. Finally, we use our molecular barcoding technique to interrogate the repopulation dynamics of a breast cancer PDX model, supporting their use as complex model systems suitable for studying the origins and consequences of tumour heterogeneity.

Declaration

This dissertation is the result of my own work, except where specific reference is made to the work of others done in collaboration.

The contents are original and have not been submitted, and are not concurrently being submitted, in part or in whole, for consideration for any other degree or diploma or other qualification at the University of Cambridge or at any other University or similar institution.

This dissertation does not exceed 60,000 words, as required by the Clinical Medicine and Veterinary Medicine Degree Committee.

John William Cassidy, MSci

September 2018

Acknowledgements

This work would not have been possible without the support offered by the Caldas lab and the CRUK Cambridge Institute over the past four years. In particular I would like to thank my supervisors, Dr Alejandra Bruna and Professor Carlos Caldas for continued support and guidance. Particular credit is due to Drs Oscar Rueda, Bernard Pereira & H. Raza Ali for continuous professional support, council and advice. Likewise, I would like to thank my fellow PhD students Rajbir Batra, Wendy Greenwood and Dr Stephen-John Sammut.

My ability to write and defend this thesis was dependent on the supportive environment created by my colleagues at Cambridge Cancer Genomics. For this, I would like to acknowledge the support of Drs Harry W Clifford, Nirmesh Patel and Hannah Thompson. I could not think of a better team with whom to share this next chapter in my life.

Lastly, I would like to thank Dr Olivia Harris & Penny Allan for continued personal support over the last few years, together with Professor James Cassidy for teaching me to strive for excellence.

This work has been supported by CRUK and the excellent core facilities at the CI, in particular Flow Cytometry, Microscopy, Histology, Genomics and the Biological Resources Unit. As ever, we are grateful to the women who donated tissues to further our research & the mice that were sacrificed in our pursuit of knowledge.

Dedicated to the memory of

PROFESSOR JAMES HILL “HAMISH” BARBER

MD FRSE FRCGP FRCP FHKCGP

Table of Contents

CHAPTER 1 - INTRODUCTION

1.0 INTRODUCTION.....	1
1.1 GENOMIC DRIVERS OF TUMOUR HETEROGENEITY	4
1.1.1 <i>Genomic Instability & Tumour Heterogeneity</i>	7
1.1.2 <i>Intertumour Heterogeneity</i>	9
1.1.3 <i>Intratumour Heterogeneity</i>	11
1.2 THE EPIGENOMIC LANDSCAPE IN CANCER	14
1.2.1 <i>Regulation of the Epigenome</i>	14
1.2.2 <i>Somatic Mutations in Epigenomic Modulators</i>	16
1.3 CONTEXT DEPENDENT SOURCES OF HETEROGENEITY	17
1.3.1 <i>Tumour Microenvironment</i>	17
1.3.2 <i>Phenotypic Plasticity</i>	18
1.3 CLINICAL IMPLICATIONS OF TUMOUR HETEROGENEITY	20
1.3.1 <i>Stratified Medicine</i>	20
1.3.2 <i>Resistance to Cancer Therapy</i>	22
1.3.3 <i>Metastatic Progression</i>	24
1.4 BREAST CANCER TUMOUR INITIATING CELLS (BC-TICs)	27
1.5 INTERROGATING TUMOUR HETEROGENEITY	33
1.5.1 <i>Genomic Clustering Techniques</i>	33
1.5.2 <i>Phenotypic Lineage Tracing Technologies</i>	34
1.5.3 <i>Origins of Drug Resistance by Lentiviral Lineage Tracing</i>	35
1.6 OESTROGEN RECEPTORS IN BREAST CANCER.....	38
1.6.1 <i>Oestrogen Receptor Structure and Function</i>	40
1.6.2 <i>ERα Signalling in Cell Fate Decisions</i>	41
1.6.3 <i>ERα as a Target of Cancer Therapies</i>	43
1.6.4 <i>Mechanisms of Resistance to Endocrine Therapies</i>	44
1.6.5 <i>Pluripotency Transcription Factors in Tamoxifen Resistance</i>	47
1.7 PRE-CLINICAL MODELS OF BREAST CANCER.....	51
1.7.1 <i>Breast Cancer Cell Lines</i>	51
1.7.2 <i>Breast Cancer Mouse Models</i>	53
1.7.3 <i>In vitro Patient Derived Models of Breast Cancer</i>	55

1.7.4 Patient Derived Xenografts.....	55
1.8 SUMMARY	60

CHAPTER 2 - MATERIALS & METHODS

2.1 ANIMAL STUDIES.....	64
2.1.1 PDX Biobank.....	65
2.1.2 Generation of Patient Derived Tumour Cells (PDTCs).....	67
2.1.3 In vivo Limiting Dilution Assay.....	67
2.1.4 Post-mortem Protocol for Identifying Metastatic Sites.....	68
2.2 CELL CULTURE ASSAYS	69
2.2.1 Mammosphere Growth Assays.....	69
2.2.2 Wound Healing Assays.....	70
2.2.3 Drug Screens and Dose Response Assays.....	71
2.2.4 Proliferation Assays.....	72
2.2.5 Matrigel 3D Growth Assays.....	72
2.2.6 In vitro Limiting Dilution Assays.....	73
2.3 CLONING, PLASMID PREPARATION & INFECTION	74
2.3.1 Cloning.....	74
2.3.2 Lentiviral Packaging	78
2.3.3 Viral Titration and Multiplicity of Infection (MOI).....	81
2.3.4 Lentiviral Infection.....	83
2.3.5 Puromycin Selection	83
2.3.6 Flow Cytometry and Cell Sorting	85
2.3.7 siRNA transfection.....	86
2.3.8 Gel Electrophoresis, extraction and purification.....	86
2.4 MOLECULAR PROFILING	87
2.4.1 Real-time quantitative polymerase chain reaction (RT-qPCR).....	87
2.4.2 Western Blotting.....	88
2.4.3 Immunofluorescence and Confocal Microscopy.....	89
2.4.4 Immunohistochemistry and Histology.....	90
2.5 BARCODE ANALYSIS	92
2.5.1 Next Generation Barcode Sequencing & Analysis.....	92

2.5.2 Barcode Composition Analysis.....	94
2.5.3 Calculation of Barcode Overlap Between Samples	94
2.5.4 Spike-in Experiments.....	95
2.6 NEXT GENERATION SEQUENCING	96
2.6.1 Whole Exome Sequencing.....	96
2.6.2 Shallow Whole Genome (sWGS).....	96
2.6.3 PyClone Clonal Reconstruction.....	97
2.7 STATISTICAL ANALYSIS.....	98

CHAPTER 3 - TRACKING TAMOXIFEN RESISTANCE WITH COMPLEX BARCODES

3.1 INTRODUCTION.....	100
3.2 CHARACTERISATION OF AN ESTABLISHED OESTROGEN-INSENSITIVE CELL LINE.....	103
3.3 SELECTION OF TAMOXIFEN RESISTANT MCF7 SUB-LINES.....	109
3.4 GENERATION OF CLONTRACER MCF7 LINE (MCF7-CT).....	115
3.5 TAMOXIFEN RESISTANCE STUDIES IN BARCODED MCF7s.....	128
3.6 BARCODE DISTRIBUTION IN 4-OHT RESISTANCE	142
3.7 ENRICHMENT OF COMMON BARCODES IN RESISTANCE	149
3.8 DYNAMICS OF RESISTANCE OVER TIME	156
3.9 GENERAL DISCUSSION	165

CHAPTER 4 – THE ROLE OF BC-TICS IN TAMOXIFEN TREATMENT & RESISTANCE

4.1 INTRODUCTION.....	173
4.2 M-ICS ARE ENRICHED BY 4-OHT TREATMENT	177
4.3 OCT4 & SOX2 IN MAMMOSPHERE FORMATION	196
4.4 GENERATION OF OCT4VEX MCF7 CELL LINE.....	203
4.5 EMT IN OCT4-VEX, TAMR AND PARENTAL MCF7s.....	213
4.6 <i>IN VIVO</i> CHARACTERISTICS OF OCT4-VEX MCF7s	221
4.7 TAMOXIFEN RESISTANT PDX MODELS.....	231
4.8 BC-TIC MARKERS IN THE METABRIC COHORT.....	242
4.9 GENERAL DISCUSSION	249

CHAPTER 5 - LINEAGE TRACING OF PDX MODELS WITH COMPLEX BARCODES

5.1 GENERAL INTRODUCTION.....	256
5.2 LENTIVIRAL INFECTION OF PDX MODELS.....	260
5.3 LOSS OF DIVERSITY IN STG282CT PDX MODELS	274
5.4 STG282-CT REPOPULATION DYNAMICS	287
5.5 PHENOTYPIC SUBTYPING OF BARCODED CLONES.....	295
5.6 PROFILING OF METASTATIC SITES IN STG282CT	308
5.7 GENERAL DISCUSSION	321

CHAPTER 6 – SUMMARY, CONCLUSIONS & FUTURE WORK

6.1 CONTEXT FOR THIS WORK: BREAST CANCER INCIDENCE & THE CLINICAL IMPORTANCE OF TUMOUR EVOLUTION.....	330
6.2 RESISTANCE TO ENDOCRINE THERAPY	332
6.3 CLONAL TRACING IN COMPLEX PRECLINICAL MODELS.....	336

APPENDIX

PUBLICATIONS RELATED TO THIS THESIS.....	340
REFERENCES	ERROR! BOOKMARK NOT DEFINED.

List of Tables & Figures

CHAPTER 1 - INTRODUCTION

FIGURE 1.1 – TUMOUR HETEROGENEITY	6
FIGURE 1.3 – INTEGRATIVE CLUSTERS DEFINED IN THE METABRIC COHORT	21
FIGURE 1.4 – UNIFIED THEORY OF TUMOUR HETEROGENEITY	32
FIGURE 1.6 - OESTROGEN AND TAMOXIFEN ACTION AT ERA	39
FIGURE 1.7 – ESTABLISHMENT OF A LARGE CLINICALLY ANNOTATED PDX COHORT	59
FIGURE 1.8 - MAINTAINING HETEROGENEITY IN PDX MODELS	62

CHAPTER 2 - MATERIALS & METHODS

TABLE 2.1.3 - CLINICAL AND MOLECULAR FEATURES OF PDX MODELS	66
FIGURE 2.3.0 – PLASMID MAPS	77
FIGURE 2.3.1 – COMPETENT E. COLI COMPARISON	79
FIGURE 2.3.2 - CLONTRACER VIRAL PRODUCTION	80
FIGURE 2.3.3 - MULTIPLICITY OF INFECTION (MOI) CALCULATIONS	82
FIGURE 2.3.4 - POLYBRENE OPTIMISATION	84
TABLE 2.4.1 – QPCR TAQMAN PROBES FROM THERMO-FISCHER	88
TABLE 2.4.4.1 – DETAILS OF PRIMARY ANTIBODIES USED IN IHC	91
TABLE 2.4.4.2 – DETAILS OF SECONDARY ANTIBODIES USED IN IHC	91
TABLE 2.5.1 – PRIMERS USED FOR BARCODE LIBRARY PREPARATION	93

CHAPTER 3 - TRACKING TAMOXIFEN RESISTANCE WITH BARCODES

FIGURE 3.2.1 – E2 & 4-OHT EFFECTS ON MCF7S & TAMRS	106
FIGURE 3.2.2 – E2 & 4-OHT EFFECTS ON MCF7 & TAMR PROLIFERATION	108
FIGURE 3.3.1 – DERIVING 4-OHT RESISTANT MCF7S	112
FIGURE 3.3.2 - SENSITIVITY OF DERIVED RESISTANT LINES TO E2 & 4-OHT	114
FIGURE 3.4.1 - CLONTRACER CELLULAR CLONE TRACKING SCHEMATIC	119
FIGURE 3.4.2 – BARCODE AMPLIFICATION BALANCE	121
FIGURE 3.4.3 - PUROMYCIN TOXICITY AND VIRAL TITRES FOR CLONTRACER IN MCF7	123
FIGURE 3.4.5 – COMPLEXITY OF BARCODES IN MCF7 POOL	127
FIGURE 3.5.1 - EXPERIMENTAL DESIGN FOR RESISTANCE SCREEN	132
FIGURE 3.5.2 - ANALYSIS PIPELINE FOR CLONTRACER BARCODES	134
FIGURE 3.5.3.1 – 4-OHT RESPONSE IN SENSITIVE LINES, BY DOSE RESPONSE	136

FIGURE 3.5.3.2 – 4-OHT RESPONSE IN RESISTANT LINES, BY DOSE RESPONSE	138
FIGURE 3.5.4 – PHENOTYPIC RESPONSE TO ESTROGEN OF RESISTANT LINES	140
FIGURE 3.5.5 – RESISTANT LINES DIFFER IN RAPAMYCIN SENSITIVITY	141
FIGURE 3.6.1 – BARCODE ENRICHMENT IN RES1 & CON1 CELL LINES AFTER 6 MONTHS	144
FIGURE 3.6.2 – RELATIVE BARCODE ABUNDANCE (RBA) IN EACH CELL LINE	146
FIGURE 3.6.3 – CUMULATIVE RBA ACROSS CELL LINES	148
FIGURE 3.7.1 – CORRELATION OF BARCODE ABUNDANCES BETWEEN REPLICATES	151
FIGURE 3.7.2.1 – SHARED BARCODES IN SENSITIVE CELL LINE REPLICATES	152
FIGURE 3.7.2.2 – SHARED BARCODES IN RESISTANT CELL LINE REPLICATES	153
FIGURE 3.7.3 – SHARED BARCODES ACROSS CELL LINES	155
FIGURE 3.8.1 – VAFS AND PYCLONE IN DEVELOPMENT OF RESISTANCE TO 4-OHT	160
FIGURE 3.8.2.1- DYNAMICS OF BARCODE ABUNDANCE IN CONTROL LINES	161
FIGURE 3.8.2.2 - DYNAMICS OF BARCODE ABUNDANCE IN RESISTANT LINES	162
FIGURE 3.8.3 - QUIESCENCE IN RESISTANT CELL LINES	164
 CHAPTER 4 – THE ROLE OF BC-TICS IN TAMOXIFEN TREATMENT & RESISTANCE	
FIGURE 4.2.1 – QPCR FOR STEM CELL MARKERS	184
FIGURE 4.2.2 – COMBINED ANALYSIS OF QPCR FOR STEM CELL MARKERS	185
FIGURE 4.2.3 – MAMMOSPHERE FORMATION ASSAYS	187
FIGURE 4.2.4 – TAMOXIFEN M-IC SELECTION	189
FIGURE 4.2.5 – OCT4 BINDING SITE IN ERA PROMOTOR	191
FIGURE 4.2.6 – MAMMOSPHERES ARE RESISTANT TO TAMOXIFEN2	193
FIGURE 4.2.7 - SCHEMATIC OF MCF7 M-IC SELECTION BY 4-OHT	194
FIGURE 4.2.8 – BARCODE CORRELATION BETWEEN M-ICS AND 4-OHT RESISTANCE	195
FIGURE 4.3.1 –SOX2/OCT4 KNOCKDOWNS	198
FIGURE 4.3.2 – ROLE OF SOX2/OCT4 IN MCF7 MFE	200
FIGURE 4.3.3 – ROLE OF SOX2/OCT4 IN TAMR MFE	202
FIGURE 4.4.1- GENERATION OF OCT4-VEX-POLY MCF7 CELL LINES	206
FIGURE 4.4.2 - CHARACTERIZATION OF OCT4-VEX-POLY MCF7S	208
FIGURE 4.4.3 - MFE IN OCT4-VEX-POLY MCF7S	210
FIGURE 4.4.4 - IN VITRO LDA OF OCT4-VEX-POLY MCF7S	212
FIGURE 4.5.1 - MAMMOSPHERE IHC	216
FIGURE 4.5.2- TAMR CELLS HAVE A MIGRATORY PHENOTYPE	218
FIGURE 4.5.3 – TAMR CELLS HAVE SOME FEATURES OF EMT	220

FIGURE 4.6.1 - CHARACTERIZATION OF TUMOUR GROWTH IN OCT4-VEX-POLY MCF7S	224
FIGURE 4.6.2 - E2 DEPENDENCE IN OCT4-VEX-POLY MCF7 TUMOURS.	226
FIGURE 4.6.3 - IHC OF OCT4-VEX-POLY MCF7 TUMOURS	228
FIGURE 4.6.4 - IN VIVO LDA OF OCT4-VEX-POLY MCF7	230
FIGURE 4.7.1 - 4-OHT RESISTANCE IN PDX MODELS	234
FIGURE 4.7.2 - QPCR FOR STEM CELL MARKERS	236
FIGURE 4.7.3 - COMBINED ANALYSIS OF QPCR FOR STEM CELL MARKERS	237
FIGURE 4.7.4 - TAMOXIFEN ALDH+ SELECTION	239
FIGURE 4.7.5 - ALDH+ SORTED PDTCS ARE RESISTANT TO 4-OHT	241
FIGURE 4.8.1 - BC-TIC PHENOTYPE IN THE METABRIC DATASET	244
FIGURE 4.8.2 - METABRIC SURVIVAL	246
FIGURE 4.8.3 - POU5F1 IN METABRIC	248
 CHAPTER 5 - LINEAGE TRACING OF PDX MODELS WITH BARCODES	
FIGURE 5.2.1 - PUROMYCIN RESISTANCE & PDTCT GATING STRATEGY	265
FIGURE 5.2.2 - STG282X3 HIV-ZSGREEN1 IMPLANTATION	267
FIGURE 5.2.3 - STG282X3 CLONTRACER INFECTION & IMPLANTATION	269
FIGURE 5.2.4 - STG282X3 CLONTRACER X0 (STG282CT-X0) HISTOLOGY	271
FIGURE 5.2.5 - STG282CT-X0 SWGS	273
FIGURE 5.3.1 - STG282CT MODEL EXPANSION.	280
FIGURE 5.3.2 - BARCODE COMPLEXITY IN STG282CT-X1 POOL	282
FIGURE 5.3.3 - ESTIMATING CELL NUMBER FROM BARCODE ABUNDANCE	284
FIGURE 5.3.4 - PASSAGE REDUCES CLONAL BARCODE COMPLEXITY	286
FIGURE 5.4.1 - STG282CT REPOPULATION DYNAMICS	291
FIGURE 5.4.1 A - STG282CT REPOPULATION DYNAMICS	292
FIGURE 5.4.1 B - STG282CT REPOPULATION DYNAMICS	293
FIGURE 5.4.1 C - STG282CT REPOPULATION DYNAMICS	294
FIGURE 5.5.1 A, B & C - STG282CT CLONAL PHENOTYPES	302
FIGURE 5.5.1 D & E - STG282CT CLONAL PHENOTYPES	303
FIGURE 5.5.2 A - CONTRIBUTION OF CLONAL PHENOTYPES TO REPOPULATION KINETICS	305
FIGURE 5.5.2 B - CONTRIBUTION OF CLONAL PHENOTYPES TO REPOPULATION KINETICS	306
FIGURE 5.5.2 C - CONTRIBUTION OF CLONAL PHENOTYPES TO REPOPULATION KINETICS	307
FIGURE 5.6.1 - IDENTIFICATION OF METASTASIS IN STG282CT	313
FIGURE 5.6.2 - CLONAL PHENOTYPES IN STG282CT METASTATIC DEPOSITS	315

FIGURE 5.6.3 – OVERLAPPING BARCODE ABUNDANCE IN METASTASES	317
FIGURE 5.6.4 – CLONAL PHENOTYPE REPRESENTATION IN METASTASES	319
FIGURE 5.6.5 – CLONE INITIATING CAPACITY (CIC) OF BARCODED CELLS	320

List of Abbreviations

#	Ataxia-telangiectasia mutated (ATM), 328	Coactivator Associated Arginine Methyltransferase 1 (CARM1), 34
11-zinc finger protein (CTCF), 41	ATP-binding cassette (ABC), 23	Colorectal Cancer (CRC), 4
2-(N-morpholino) ethanesulfonic acid (MES), 82	B	Copy number (CN), 1
3,3-diaminobenzidine tetrahydrochloride (DAB), 83	Bcl-2-like protein 11 (BIM), 32	Copy number aberration (CNA), 5
4',6-diamidino-2- phenylindole (DAPI), 65	Biological Resources Unit (BRU), 57	Cytochrome P450 2D6 variant system (CYP2D6), 64
4-(2-hydroxyethyl)-1- piperazineethane sulfonic acid (HEPES), 60	Bovine Serum Albumin (BSA), 60	D
5-aza-2-deoxycytidine (AZA), 39	C	Dimethyl sulfoxide (DMSO), 58
5-Azacytidine (5-Aza), 44	Cancer Associated Fibroblast (CAF), 250	Disability Adjusted Life Year (DALY), 322
A	Cancer Stem Cell (CSC), 2	DNA methyltransferase-1 (DNMT-1), 39
Activation Function (AF), 33	Catalogue of Somatic Mutations in Cancer (COSMIC), 42	DNA-binding domain (DBD), 33
Adenomatous polyposis coli (APC), 4	Charcoal-stripped serum (c/s serum), 102	Dulbecco's Modified Eagle Medium (DMEM), 60
Aldehyde dehydrogenase (ALDH), 22	Chromatin immunoprecipitation followed by high- throughput sequencing (ChIP-seq), 162	E
Alpha-6 integrin (CD49f), 22	Chronic Myeloid Leukaemia (CML), 31	E26 Transformation-Specific (ETS)
Analysis of Variance (ANOVA), 91	Clone-Initiating Cell (CIC), 270	Enzyme-Linked Immunosorbent Assay (ELISA), 74
Area under the curve (AUC), 224		Epidermal Growth Factor (EGF), 62
Aromatase Inhibitor (AI), 37		

Epidermal Growth Factor
 Receptor (EGFR), 17
 Epithelial to Mesenchymal
 Transition (EMT), 14
 ER α -positive (ER α +), 93
 ER α -negative (ER α -), 39
Escherichia coli (*E. coli*), 67
 Ethanol (EtOH), 83
 Ethylenediaminetetraacetic
 acid (EDTA), 60
 Extracellular Matrix (ECM),
 13
 Extreme Limiting Dilution
 Analysis (ELDA), 61

F

Fluorescein Isothiocyanate
 (FITC), 78
 Fluorescence Activated Cell
 Sorting (FACS), 254
 Fluorescent *in situ*
 Hybridisation (FISH), 6
 Foetal Bovine Serum (FBS),
 58
 Forkhead Box A1 (FOXA1),
 36
 Forward scatter (FSC), 253

G

GATA-binding protein 3
 (GATA3), 36
 Genetically Engineered
 Mouse (GEM), 47

Glyceraldehyde 3-phosphate
 dehydrogenase (GAPDH),
 80
 Growth Factor Reduced
 (GFR), 65
 Growth Regulating Estrogen
 Receptor Binding 1
 (GREB1), 44

H

Hank's Balanced Salt
 Solution (HBSS), 60
 Hexadimethrine bromide
 (Polybrene), 76
 High Throughput Screen
 (HTS), 224
 Histone deacetylase (HDAC),
 39
 Histone deacetylase 3
 (HDAC3), 34
 Homeobox transcription
 factor Nanog-delta 48
 (NANOG), 42
 Horseradish peroxidase
 (HRP), 83
 Human Epidermal Growth
 Factor Receptor 2 (HER2),
 7

I

Immunohistochemistry
 (IHC), 83
 Induced Pluripotent Stem
 Cell (iPSC), 44

Inhibitory Concentration.
 50% (IC50), 96
 Insulin-like Growth Factor-1
 (IGF-1), 40
 Integrative cluster (IC), 235
 Integrative Clusters
 (IntClusters), 245
 Integrative Genomics
 Browser (IGV), 183
 Interleukin-6 (IL-6), 162
 Intestinal Stem Cell (ISC), 48

J

Janus kinase 2 (JAK2), 8
 Jun D proto-oncogene
 (JUND), 15

K

Kruppel-like Factor 4
 (KLF4), 44

L

Leucine-rich-repeat
 containing G-Protein
 Coupled Receptor 5
 (LGR5), 48
 Limiting Dilution
 Transplantation Assay
 (LDA), 22
 Long Terminal Repeat (LTR),
 109
 Luria Broth (LB), 67

Lysine (K)-specific
demethylase 6A (KDM6A),
29

M

Major Histocompatibility
Complex (MHC), 253
Mammalian Target of
Rapamycin (mTOR), 40
Mammary Epithelial Growth
Media (MEGM), 60
Mammosphere Forming
Assay (MFA), 22
Mammosphere Forming
Efficacy (MFE), 23
Mammosphere Initiating
Cells (M-IC), 170
Mammosphere Media (MM),
62
MCF7 ClonTracer (MCF7-
CT), 231
Metastasis-Associated
Macrophage (MAM), 47
Microsatellite Instability
(MSI), 5
Mitogen-activated Protein
Kinase (MAPK), 39
Mitogen-activated Protein
Kinase Kinase Kinase 1
(MAP3K1), 29
Molecular Taxonomy of
Breast Cancer
International Consortium
(METABRIC), 91

Mouse Mammary Tumour
Virus (MMTV), 47
Multiplicity of Infection
(MOI), 74
MutL homolog 1 (MLH1), 8
Myeloproliferative
Neoplasms (MPNs), 8

N

Natural Killer (NK), 14
N-Desmethyltamoxifen
(NDMTAM), 64
Next Generation Sequencing
(NGS), 31
No Supplement (NS), 106
NOD.Cg-*Prkdc^{scid}*
IL2rg^{tm1Wjl}/SzJ (NSG), 49
NOD.Cg-*Rag1^{tm1Mom}*
IL2rg^{tm1Wjl}/SzJ (NRG), 49
Non-Small Cell Lung Cancer
(NSCLC), 31
Nuclear Receptor (NR), 33
Nuclear Receptor
Corepressor 1 (N-CoR1),
34
Nuclear Receptor
Corepressor 2 (N-CoR2),
34

O

Octamer-binding
Transcription Factor 4
(OCT4), 42
Oestrogen Receptor (ER), 1

Oestrogen Response Element
(ERE), 173

P

Paired box gene 2 (PAX2), 41
Patient Derived Tumour Cell
(PDTc), 60
Patient Derived Tumour
Xenograft (PDX), 3
Penicillin-Streptomycin
(PenStrep), 62
Phosphatidylinositol 3-
kinase
(PI3K), 8
Phosphatidylinositol 3-
kinase alpha catalytic
subunit (PIK3CA), 8
Phospho-Buffered Saline
(PBS), 71
Phosphoglycerate kinase
(PGK), 196
Poly ADP Ribose Polymerase
(PARP), 17
Polycomb group RING finger
protein 4 (BMI1), 24
Polycomb Repressive
Complex 2 (PRC2), 44
Polymerase Chain Reaction
(PCR), 32
Programmed-Death Ligand 1
(PD-L1), 14
Protein C Receptor (PROCR),
22

Q

Quality Control (QC), 122

R

Red Florescent Protein

(RFP), 253

Regulatory T-cell (T_{reg}), 14

Relative Barcode Abundance

(RBA), 135

RNA polymerase II (RNAPII),
173

RNA sequencing (RNA-seq),
31

S

Selective Oestrogen Receptor
Modulator (SERM), 36

Selective Oestrogen Receptor
Degradar (SERD), 37

Sex determining region Y box
2 (SOX2), 42

Shallow Whole Genome
Sequencing (sWGS), 255

Shannon Diversity Index
(SDI), 135

Short-Interfering RNA
(siRNA), 79

Side scatter (SSC), 253

Single Nucleotide variant
(SNV), 5

Sociodemographic Index
(SDI), 322

Sodium Dodecyl Sulphate
(SDS), 82

STG282 ClonTracer

(STG282CT), 342

Suberoylanilide hydroxamic
acid (SAHA), 39

T

Terminal End-Bud (TEB), 36,
167

Telomerase Reverse

Transcriptase (TERT), 65

Tet methylcytosine
dioxygenase 2 (TET2), 8

The Cancer Genome Atlas
(TCGA), 91

Transformed human
mammary epithelial
(HMLER), 24

Transforming Growth Factor
Beta ($TGF\beta$), 47

Transforming Growth Factor
Beta Receptor 3
($TGFBR3$), 15

Transforming Units (TU), 74

Trichostatin A (TSA), 39

Triple Negative Breast
Cancer (TNBC), 9

Tris base, acetic acid and
EDTA (TAE), 68

Tumour Associated
Macrophage (TAM), 250

Tumour Initiating Cell (TIC),
22

Tumour Microenvironment
(TME), 13

Tumour Necrosis Factor
Alpha ($TNF-\alpha$), 162

U

Ubiquitin carboxyl-terminal
hydrolase (BAP1), 29

Ultraviolet Radiation (UV), 5

V

Variant Allele Frequency
(VAF), 95

W

Woodchuck Hepatitis Virus
Post-transcriptional
Regulatory Element
(WPRE), 109

Y

Years Lost due to Disability
(YLD), 322

Years of Life Lost (YLL), 322

Z

Z-4-Hydroxytamoxifen (4-
OHT), 64

Zinc finger E-Box1 (ZEB1),
24

 β

β 2-Microglobulin (B2M), 8

[THIS PAGE IS INTENTIONALLY BLANK]

Chapter 1 - Introduction

1.0 Introduction

Cancer has been known to be heterogeneous since its detailed study by experimental pathologists began at the start of the 19th century. At first, differences in cellular morphology were described (1), followed by heterogeneity across surface marker expression (2) and later differences in tumour growth rates (3) and response to therapy (4). Recently, large and small scale profiling endeavours have helped elucidate the true scale of diversity across human neoplasms (5). Early work in breast cancer, for example, allowed stratification of patients based on the presence of oestrogen receptor alpha (ER α), which led to the successful targeting of tamoxifen for ER α positive (ER α +) patients (6). More recent work has enabled comprehensive stratification of breast and other cancers (5,7). For example, in breast cancer, a 50 gene signature (PAM50) can be used to stratify patients into four intrinsic subtypes (luminal A, luminal B, HER2-enriched & basal-like) with distinct clinical outcomes (8,9). Our lab has recently integrated copy number (CN) data with transcriptomics to improve on this classification and uncover 11 distinct Integrative Clusters of breast cancer (7). Defined in 2,000 tumours, this classification was validated in over 7,500 tumours (10) and shown to

clearly associate with discrete clinical outcomes, suggesting these clusters represent distinct biological entities. Improving the taxonomy of cancer is the initial step towards a better understanding of the drivers of tumour growth and consequently towards improved precision medicines. It is hoped that this strategy may ultimately lead to development of the next generation of targeted therapies (11).

Further to heterogeneity between patients, it has long been known that tumours harbour distinct cellular populations within their bulk (1,12). For example Fidler and Kripke reported in 1977 that clonal populations derived from mouse metastatic melanomas varied extensively in their ability to seed metastasis in syngeneic hosts (13). Of note amongst the relatively early research into tumour heterogeneity, were the important observations using human patients, although highly unethical by today's standards these offered unequivocal evidence for functional heterogeneity and fuelled five decades of subsequent research. For example, various studies around the late 1960s showed by *in vivo* radiolabelling that the morphologically distinguishable populations of human leukemic cells differed remarkably in their proliferative potential (14–16). The observation that human cancers contain functionally different populations was echoed by *Southam et al.*, who in 1962 showed that autologous engrafted human tumour cells differ in their ability to reform tumours (17).

These early studies showed that tumours are not simply a growth of homogeneous cells with equal proliferative potential and tumour forming ability, but a heterogeneous mixture of cellular populations. The observation that tumour cells differ in their ability to xeno- and auto-transplant was added to by seminal studies on teratocarcinomas (18), small cell lung carcinomas (19) and mammary adenocarcinomas (20), to give rise to the cancer stem cell (CSC) model of tumour development and heterogeneity. Early evidence that genetic aberrations were the cause of a tumours phenotypic traits (21) supported the idea that somatic evolution of genomic clones could occur, allowing Darwinian selection in response to spatial and temporal selective pressures (12). Together these

theories have influenced a significant proportion of the cancer research occurring today. However, translation of this research into the clinical setting has been slower; hampered by the realisation that traditional pre-clinical models lack the heterogeneity of human malignancies (22). Recently, patient derived tumour xenografts (PDXs) have emerged as powerful preclinical models capable of recapitulating most of their originating tumour's heterogeneity (23). As such, studies on PDXs are showing strong basic and translational impact in the field, which is crucial in order to avoid the unsustainable rates of attrition in oncological drug development.

Although often considered disparate, the causes and consequences of *intertumour* heterogeneity (*i.e.* different sub-classifications of tumours from the same originating organ) and *intratumour* heterogeneity (*i.e.* different sub-clonal architectures) are inherently linked. Moreover, compelling models of tumour heterogeneity, linking the competing theories of CSCs and somatic evolution have recently been proposed (24). As computational and preclinical models improve, and as high-throughput single cell studies become mainstream, we will begin a new chapter in our understanding of the causes and consequences of tumour heterogeneity. It is hoped that this new *era* of research will lead to new insights into the complex ecosystem that is cancer and that these insights can be translated into the clinic to accurately stratify tumours and, ultimately, improve clinical outcome.

1.1 Genomic Drivers of Tumour Heterogeneity

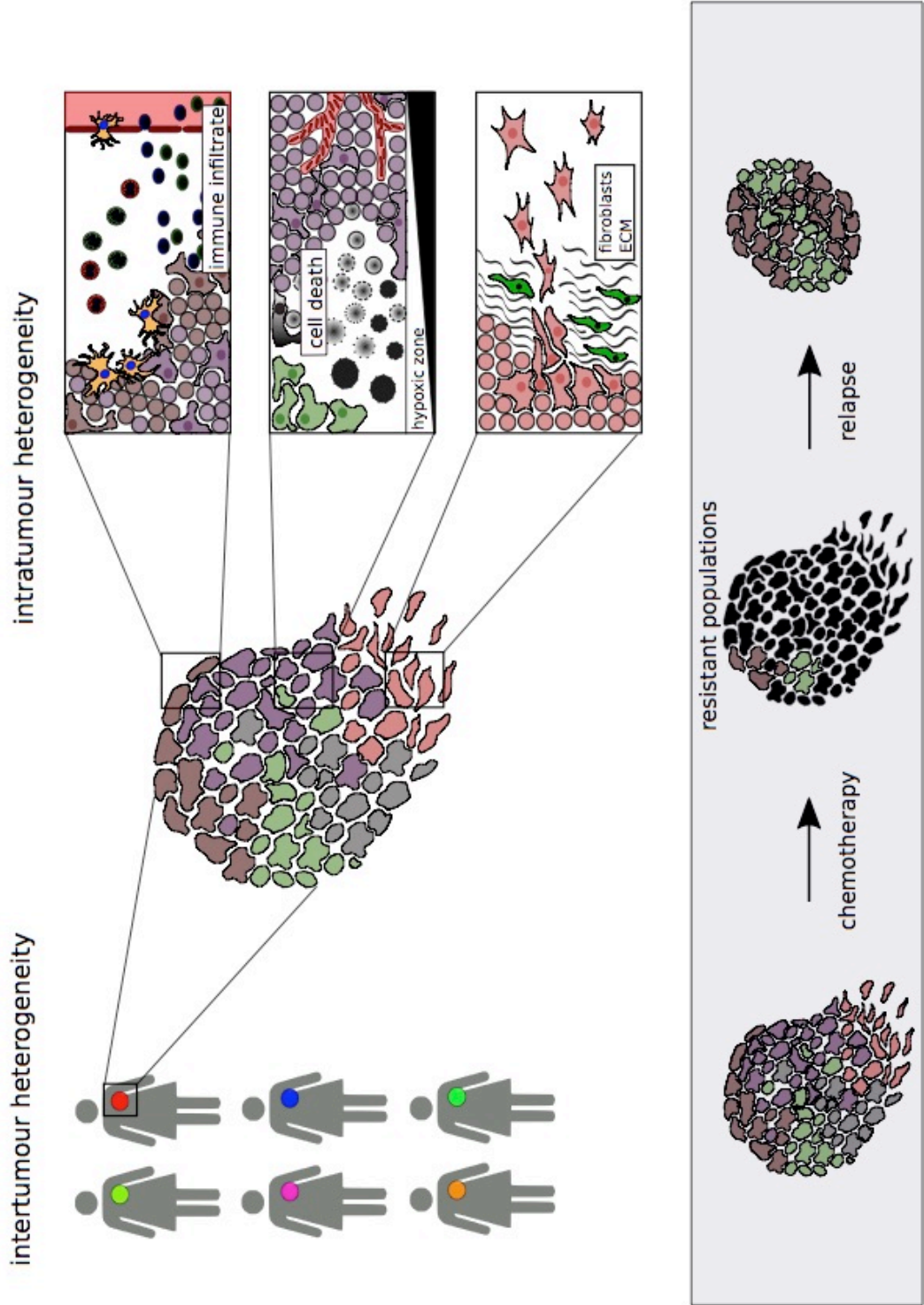
Cancer is, first and foremost, a disease of the genome. Indeed, both inter- and intratumour heterogeneity can be explained by the genomic instability inherent to a tumour's biology & the sequential acquisition of driver mutations. For example *Adenomatous polyposis coli* (*APC*) loss in colorectal cancer (CRC)) which fuels the first stages of clonal expansion (25,26). Though changes in a tumour's microenvironment (for example increase inflammation or immune cell infiltrate) or epigenetic regulation (for example *MLH1* promotor methylation in microsatellite unstable CRC) are undoubtedly required to transform a clonal expansion of benign cells into a malignancy (21, 23, 70, 77), this section will focus on the aforementioned genomic drivers.

Through the course of tumour initiation and progression, cancer cells undergo repeated mutational events, which may or may not confer a survival advantage ('fitness') on their progeny. With time, this process generates a dominant clone that will expand and dominate the site where it was generated through Darwinian selection in response to spatial and temporal selective pressures (12). When clones arise with an increased fitness (or when selective pressures change), less advantaged clones will either disappear or will be maintained as sub-clones alongside the dominant clone, acting as a reservoir from which evolution can continue (27). This compelling theory was first put forward by Peter Nowell in 1976 (12) and supported by early evidence that genetic aberrations were the cause of a tumours phenotypic traits (21) and more recent genomics research (28,29). It is now accepted that tumours harbour various layers of genomic complexity and the resultant heterogeneity can have profound effects on disease progression. Moreover, genomic instability, which fuels the diversity essential for any Darwinian process is intertwined with both the development and maintenance of tumour heterogeneity, and the clinical consequences thereof.

Figure 1.1 – Tumour Heterogeneity

*Figure 1.1 shows an overview of tumour heterogeneity. From left to right; intertumour heterogeneity ensures that no two malignancies are the same; genomic clonal populations exist within a tumour (coloured) and functional heterogeneity (shape) exists within isogeneic populations. This non-genomic intratumour heterogeneity is due to intrinsic epigenetic differences (not shown), interaction with the immune infiltrate (top panel), differences in tumour metabolism (e.g. hypoxia; middle panel) and interaction with the extracellular matrix and stromal component (bottom panel). Each of the depicted environments could have different effects on cellular functions. Highlighting functional consequences of heterogeneity, in the lower half, two genomic clones are depicted as resistant to chemotherapy and able to repopulate the tumour after treatment. Dead cells are coloured black in the centre of the lower panel. These resistant clones could be genomically distinct or isogenic but functionally distinct. Adapted from (11). **Figure overleaf.***

Figure 1.1 – Tumour Heterogeneity



1.1.1 Genomic Instability & Tumour Heterogeneity

According to Hanahan and Weinberg, genomic instability is an enabling characteristic that helps generate the hallmarks of cancer, and is the major driving force behind intra- and intertumoral heterogeneity (30,31). Throughout the process of tumour development, cancer cells can accrue thousands of mutations, some of which can even involve the gain or loss of entire chromosome arms (32). However, there is evidence that the number of mutations cannot increase endlessly without adversely affecting cell fitness (33–37), implying the existence of a limit of tolerance. Hence, cancer cells must exist in constant balance between instability-driven cell growth and the point where the consequences of gross genomic changes become lethal to the cell (32).

Solid tumours can be classified based on the dominance of single nucleotide variants (SNVs), *i.e.* M-class tumours, or copy number aberrations (CNAs), *i.e.* C-class tumours (38). The contribution of each class of somatic mutation to genomic instability, the balance between instability-driven cell growth & cell death, and tumour heterogeneity, is varied. For example, the total burden of SNVs at a given time depends on the efficacy of SNV appearance and clearance by, for example, p53-induced apoptosis (32). In some cancers, *e.g.* microsatellite instability (MSI) high CRCs, SNV burden is increased by a loss of DNA mismatch repair and this is linked to a favourable outcome. Intolerance to a high SNV burden could be due to increased immune clearance. Indeed MSI high CRCs and metastatic melanomas¹ have both been shown to respond well to immune checkpoint inhibitors (39–43).

Simulation studies have provided insight into how natural selection adjusts mutation rate in tumours with a high SNV burden (34,44). One such study found that under fluctuating environmental conditions (*e.g.* oxygen availability or temperature), the rate of SNV accumulation increased linearly until reaching a critical limit. SNV accumulation

¹ The high mutation burden in metastatic melanoma is linked to prolonged exposure to ultraviolet (UV) light and the resultant C>T transitions (76).

at a rate above this variable threshold led to population level extinction events (45). Another simulation study found that a high SNV accumulation rate initially leads to rapid tumour growth, but that beyond a certain threshold, leads to negative clonal selection and is consequently less favourable for cellular expansion (46). Interestingly, concepts such as immune surveillance, which limit the proportion of cells with a high SNV burden² are not easy to adapt to explain the toxicity associated with a high *rate* of SNV formation. However, there does exist a theoretical framework, seen in the development of life, for a limiting threshold in DNA replication error rate (47). Specifically, it has been proposed that if a SNV error rate were to exceed some catastrophic threshold, then the information in the genome would be effectively decayed and the fidelity of genome maintenance across generations would be severely impacted (34,48). This conceptual framework is supported by theoretical models and evolutionary experiments (33–36,49), applicable to unicellular organisms, multicellular organisms, and neoplastic cells (50,51). Mechanistically, it is thought that SNV error rate limits are supported by ‘gatekeeper’ genes (such as *TP53*) which may be induced by oxidative stress and/or high SNV mutation rates, halting the cell cycle and initiating apoptosis (52–55).

CNAs differ from SNVs in that they can encompass a vast region of a cell’s genome. CNAs are thought to confer substantial phenotypic plasticity, through gene duplication or deletion, and have been described as the driving force of genetic diversification (19). Indeed, there is evidence supporting the more central role of CNAs than SNVs in the development and maintenance of neoplastic cell population diversity (56–58). However, CNAs appear limited by a similar overall rate and burden limit (32,59,60). For example, in one study, fluorescent *in situ* hybridisation (FISH) of the centromeres of chromosomes 2 and 15 was used to define CNA driven genomic instability in ER α - breast cancer patients³. The authors found that cases where 45% of tumour cells had chromosomal abnormalities had a significantly better prognosis than cases with lower number of

² If we consider that high SNV burden translates to an increased frequency of neoantigen formation.

³ Chromosomes 2 and 15 were chosen due to their relatively rare specific alteration frequency in breast cancer, the authors reasoned CNAs in these chromosomes would be indicative of overall instability (7).

chromosomal abnormalities (60). Similarly, TNBCs with a gene expression signature associated with high chromosomal instability are associated with increased time until relapse compared with those with low predicted instability (59). Limits on CNA abundance could be explained by biophysical constraints (*e.g.* chromosome size limiting alignment to the centre of the nucleus and therefore metaphase efficiency) (61), gene dosage (*e.g.* amplification of neoantigens or tumour suppressors in a duplicated region) (62) or apoptosis initiated by DNA damage (*e.g.* DNA double-strand breaks initiating p53-dependent signal transduction) (63,64).

In summary, genomic instability is a major driver of tumour heterogeneity, yet whilst heterogeneity may be associated with a poor prognosis, instability itself may be associated with better patient outcome. These observations can be reconciled by considering that a high CNA burden may result from multiple clones with low levels of CNA burden or of a few clones with high levels of CNA burden (32). When CNA burden is spread among many clones, the associated prognosis is less favourable, indicating that it is the CNA burden per clone that limits tumour viability. Thus, genomic instability gives rise to heterogeneity and a polyclonal tumour, but an overly high CNA or SNV burden in any specific clone, limits its viability. Genomic instability and tumour heterogeneity is best considered as a delicate balance between favourable growth characteristics and cellular toxicity (32).

1.1.2 Intertumour Heterogeneity

The earliest events in a tumour's evolution are fuelled by specific genomic aberrations, which can have profound effects on intertumour heterogeneity. For example *ESR1* and *ERBB2* amplification, leading respectively to ER α (65) and human epidermal growth factor receptor 2 (HER2) (66) protein upregulation, can be early events in breast cancer

initiation⁴. These events might be prognostic and predictive of drug responses, suggesting they can be used to classify cancer into different subtypes. For example, ER α + tumours (80% of breast cancers) tend to have a better prognosis and are treated with oestrogen receptor antagonists (*e.g.* tamoxifen) or aromatase inhibitors (*e.g.* anastrozole), whereas HER2+ tumours (20% of breast cancers) are generally faster growing, more aggressive and are treated with antibodies⁵ against HER2 (*e.g.* trastuzumab) (67). *ESR1* and *ERBB2* amplification can also occur in the same tumours, with or without the presence of progesterone receptor (PR). Indeed, the first molecular based classification that dramatically changed clinical practice and breast cancer patient outcome was based on ER α , PR and HER2 status (67). Continued technological advances have made clear that a wide range of (epi)genomic aberrations can drive the tumorigenic process. Recently, a driver-based taxonomy of breast cancer has been defined based on copy number and gene expression data (7,10). The 11 molecular subtypes identified show distinct prognosis and molecular drivers, reaffirming breast cancer heterogeneity.

Beyond the effects of individual genomic aberrations, the order in which cells acquire mutations can have profound effects on intertumour heterogeneity and disease progression. For example, in Philadelphia chromosome negative myeloproliferative neoplasms (MPNs), recent work has demonstrated that within patients harbouring both a *Janus kinase 2* (*JAK2*) and *Tet methylcytosine dioxygenase 2* (*TET2*) mutation, those who acquired the *TET2* mutation first were less likely to present with the MPN subtype *polycythemia vera* than with *essential thrombocythemia*⁶ (68). Thus, complete phenotypic heterogeneity is observed between patients with the same mutational load depending on the order of mutational events.

⁴ It is worth noting that somatic alterations in the genome are likely not enough to drive tumour initiation in the absence of mitigating factors such as local inflammation or immune cell infiltrate.

⁵ Small molecular inhibitors of the internal tyrosine kinase domains of HER2 (and other receptor tyrosine kinases) are often used as an alternative to antibodies targeting external domains. In breast cancer, the dual inhibitor of HER2 and EGFR kinase domains, lapatinib, is approved for HER2-positive breast cancer.

⁶ The author's explanation was that *TET* mutations caused differential DNA methylation signatures, affecting target transcriptional programs of oncogenic JAK^{V617F}.

Significantly, some of these early oncogenic driver events can also shape the subsequent clonal evolution that heavily influences intratumour heterogeneity. Taking the most extreme example, hypermethylation or mutation of *MutL homolog 1 (MLH1)* leads to a hypermutator phenotype in colorectal cancer (69). This microsatellite instability phenotype both distinguishes *MLH1* mutant tumours from other CRCs and leads to widespread intratumour heterogeneity (70), which has been linked to higher resistance to therapy. More recently, mutant *phosphatidylinositol 3-kinase alpha catalytic subunit (PIK3CA)* has been shown to enable plasticity in differentiated breast cells, paving the way towards functional intratumour heterogeneity in breast cancers with *PIK3CA* mutations (71,72). More broadly, *PIK3CA* or other members of the phosphatidylinositol 3-kinase (PI3K) pathway are amongst the most commonly mutated in breast cancer (73) and cross-talk between signalling networks emanating from mutant *PIK3CA* and ER α have been shown to impact significantly on breast cancer initiation and progression (74,75).

1.1.3 Intratumour Heterogeneity

The acquisition and order of driver mutations can have profound implications for intertumour heterogeneity. However, tumours are characterised by continuous clonal evolution as they develop (76). Progeny of founder clones undergo repeated mutational events that may confer a fitness advantage with regards specific spatial or temporal selective pressures (12). Clonal evolution is present in precancerous and advanced lesions and helps define both inter- and, in particular, intratumour heterogeneity. For example, sequencing data on 234 biopsies of normal skin from four individuals showed multiple cancer-associated genes were under positive selection even in normal tissue (77). The authors observed clonal expansion of skin cells with early driver mutations across patients and overall found driver mutations at a density of 140 per cm² of sun-exposed skin. As this study focussed on pre-cancerous tissues, we cannot know if any of these early lesions would lead to tumour growth but the authors do present strong

evidence that clonal evolution occurs even in the earliest stages of a neoplasm's development (77).

Numerous groups have been able to reconstruct the clonal hierarchy of individual tumours. For example *Nik-Zainal et al.*, were able to combine deep sequencing with novel bioinformatics tools to reconstruct the clonal history of 21 breast cancers (28). The authors showed that breast cancers evolve through the infrequent acquisition of driver mutations; each of which allows clonal expansion and eventual dominance. Interestingly, as the most recent common ancestor appeared relatively early, minor clones were able to coexist and diversify alongside the dominant clone (28). This model of branched evolution allows for a genetic pool of minor clones able to fuel new stages of clonal evolution if selective pressures change. In agreement with this study, by reconstructing the clonal composition of 104 triple negative breast cancer (TNBC), our group observed a complete spectrum of molecular and clonal compositions at patient diagnosis (78).

Alongside this model of branched evolution, the survival of multiple sub-dominant clones can be explained by the spatial segregation of clones across the tumour as a whole. This pattern was hinted at in the pre-cancerous clonal expansions of the normal skin (77) and fully considered in renal cell carcinoma (79) and lung cancers (80). Indeed, in a recent study from *Caravagna et al.*, the temporal order of some genomic changes in a tumour could be inferred from multiregional sequencing (81). The authors used transfer learning⁷ to transition neural networks trained on their own datasets to large multi-region sequencing datasets from lung, breast, renal, and colorectal cancer, in each case detecting repeated evolutionary trajectories in subgroups of patient. A public release of the author's software package, 'REVOLVER', could empower researchers to stratify patient groups based on the basis of how their tumour evolved (81).

⁷ Transfer learning is a strategy in deep learning whereby hyperparameters from a neural network trained on a well-labelled dataset may be used to tune the hyperparameters of a future convolutional network applied to a similar dataset or problem (475).

Furthermore, the notion that some clonal populations may cooperate through reciprocal signalling pathways can further contribute to tumour heterogeneity (82). In this model, the survival of a minor clone can be assured if it contributes some degree of paracrine signalling to the dominant clone. Ensuring expansion of the latter will never come at the expense of the former. Interestingly, this model does not require that cellular populations are genomically distinct, indeed functional heterogeneity has been observed even in isogenic populations (83).

1.2 The Epigenomic Landscape in Cancer

Like genomic mutations, epigenetic modifications are characteristic of all tumours (84). Indeed, it has been shown that the somatic mutation burden is so inextricably linked to cell-of-origin specific chromatin accessibility and modification, that, together with replication timing, this can explain up to 86% of the variance in mutation rates across a cancer genome (85). Additionally, genomic heterogeneity itself cannot explain the complete phenotypic diversity of tumours: even genetically homogenous populations of cancer cells show remarkable diversity in their response to therapeutic selective pressure and other sources of environmental stimuli (83,86). This observation suggests profound epigenomic heterogeneity may play a clinically important role in tumour biology.

1.2.1 Regulation of the Epigenome

Epigenetics is the study of heritable mechanisms of phenotypic regulation independent of genetic variation. As the vast majority of cells in an adult organ are genetically identical, cell type heterogeneity and the control of cell fate during development, are attributed to epigenomic regulators (87). On the molecular level, mechanisms of developmental control are frequently co-opted by tumours during their development, these include: post-translational modification of histones; histone variants; DNA methylation; differential utilisation of non-coding RNAs as well as chromatin remodelling (88). Indeed, the idea that such epigenetic changes may be direct drivers of tumour progression, rather than simply passenger events, was first proposed by Robin Holliday in 1979 and has since been shown in ovarian cancer (89) and B-cell lymphomas (90).

In particular, two classes of epigenomic regulators, DNA methylation and histone tail modification, have been extensively studied as contributors to tumour heterogeneity (91). Each is thought to contribute to transcriptional regulation through control of

functional genomic boundaries between open (euchromatin) and tightly packed (heterochromatin) DNA (92). The blurring of euchromatin and heterochromatin boundaries during tumour progression is thought to lead to more global restructuring of the genomic architecture (93,94) and potentially to specific cell-to-cell phenotypic variability (92). Moreover, as both DNA methylation and histone tail modification are enzymatically controlled processes, it has been suggested that they may be differentially inherited by daughter cells, further leading to phenotypic plasticity across the tumour and, ultimately, tumour heterogeneity (93).

More generally, epigenetic regulation is tightly associated to a cell's transcriptional activity (95) and consequently may influence its reaction to external stimuli, including cancer therapies (96–98). Aside from transcriptional control, profound heterogeneity is seen in the post-transcriptional landscape in breast cancer (99–101). By integrating mRNA and microRNA expression data from 1,302 breast tumours, our group was able to identify multiple prognostic microRNA expression signatures (99). These prognostic signatures were associated with immune infiltrate and could only be externally validated in CNA-devoid breast cancer subtypes, suggesting a particular role of microRNAs in the progression of this breast cancer subtype (99). Interestingly expression of specific microRNAs has also been associated with clinically relevant phenotypes, such as the progression of epithelial to mesenchymal (EMT) (102,103) and the development of drug resistance (104).

In summary, the study of epigenetic intratumor heterogeneity is necessary for a holistic understanding of clonal evolution and therapy resistance. By understanding heterogeneity in epigenetic regulatory mechanisms, we may gain a greater understanding of overall tumour heterogeneity.

1.2.2 Somatic Mutations in Epigenomic Modulators

Mutations in epigenetic writers, readers, and erasers as well as members of chromatin-remodelling complexes, are pervasive in multiple specific cancer types, with more than 30 frequently mutated genes⁸ known to play an active role in dysregulation of the epigenome (84,105). Those most frequently mutated include histones themselves; histone acetyltransferases; histone deacetylases; histone methyltransferases; histone demethylases; DNA methyltransferases and chromatin remodelling factors (84). Interestingly, mutations in epigenetic regulators are particularly common in cancers that relapse or that are otherwise resistant to therapy. For example in ovarian clear cell carcinoma, where mutations in the chromatin modulator *ARID1A* have been seen in up to 57% of cases, but are not at all found in the less aggressive high-grade serous ovarian carcinoma subtype (106,107).

Genomic mutations occur in noncoding regulatory regions of the genome at nearly double the frequency of protein coding regions (108). Such mutations have been identified in multiple promoters and enhancer elements across a range of cancer types and stages (109,110). For example, mutations in the promotor region of *telomerase reverse transcriptase (TERT)* are seen in more than 70% of melanomas (111,112). Interestingly, mutations within the coding region of *TERT* are uncommon, but those in promotor regions are thought to increase *TERT* expression by creating a *de novo* binding motif for E26 transformation-specific (ETS) transcription factors. In breast cancer, however, a recent study found *TERT* promotor mutations⁹ in only 0.9% of the 319 cases analysed (113), reflecting intertumour heterogeneity across somatic mutations in epigenetic regulators.

⁸ If we define common mutations as those included in the catalogue of somatic mutations in cancer (COSMIC) database.

⁹ Although the authors focussed only on the hotspot mutations C228T and C250T (113).

1.3 Context Dependent Sources of Heterogeneity

In addition to genomic and heritable epigenomic diversity, environmental and non-cell autonomous sources of epigenomic variation may play a role in tumour heterogeneity. Indeed, cells within a tumour are characterised by functional heterogeneity related to their specific tumour microenvironment (TME). Components of the TME (encompassing immune infiltrate, stromal compartment and extracellular matrix (ECM)) vary significantly between tumours and contribute to spatial heterogeneity in individual neoplasms. Moreover, localised signalling from stromal and immune cells could lead to phenotypic plasticity within isogenic populations. For these reasons, we should consider a tumour as a complex ecosystem functioning at a level far exceeding the some of its parts (29).

1.3.1 Tumour Microenvironment

Tumour development depends on the co-evolution of neoplastic cells along with ECM, stromal compartment, tumour vasculature and immune cells. These tumour extrinsic factors collectively comprise the TME and are required for successful outgrowth of tumours and eventual metastases (114). The dynamic topology of the tumour, together with its microenvironment, varies drastically both between, and within the same lesions (114). Thus, the individual compartments of the TME have profound influence on both inter- and intratumour heterogeneity. Perhaps the most relevant facet of the TME for tumour stratification, given the success of immunotherapies, is the tumour's immune infiltrate (43,115). For example, Th1 and cytotoxic (CD8+) T-cell content was recently found to be the strongest prognostic factor in overall survival at all stages of CRC (116). Other studies have highlighted the vast heterogeneity seen in immune infiltrates of CRC patients (117). As such, stratification of CRCs based on immune infiltrate could both provide more accurate prognosis information, and potentially identify patients who would benefit most from immunotherapy.

In breast cancer, efforts to develop immune based subclasses are already underway. One study used immunohistochemical grading of CD8+ T-cells, regulatory T-cells (T_{regs}) and Natural Killer (NK) cells to define three breast cancer subtypes (118). Analysis of training ($n=440$) and validation ($n=382$) cohorts confirmed these subtypes had significant prognostic value for relapse free and overall survival (118). Moreover, our own studies have shown *Programmed-death ligand 1* (*PD-L1*) expression is correlated with infiltrating lymphocytes and is highly enriched in basal-like breast tumours (119). Anti-PD-L1 antibodies have shown efficacy in early clinical trials for metastatic breast cancer; it seems likely that stratification of breast cancers based on immune infiltrate or *PD-L1* expression will aid in improving the efficacy of such agents (120). Moreover, integration of immune and genomic-based classifications will further increase our resolution of intertumour heterogeneity and allow a more rational stratification of patients.

1.3.2 Phenotypic Plasticity

The epigenomic landscape can be conceptualised as containing ‘hills’ (unstable states) and ‘valleys’ (stable states). A pluripotent cell begins its differentiation process at the peak of a hill and progresses into a stable state (the valley) as it becomes restricted to a specific lineage (121). Directing the cell down its lineage-restricted path are a multitude of environmental (*e.g.* topology of the ECM (122)) and biochemical cues (autocrine and paracrine signalling) which directly stimulate the transition from one epigenetic state to the next (123). During tumour development, genomic instability, the acquisition of numerous passenger mutations, coupled with irregular activation of signalling pathways, contributes to instability in these epigenetic ‘attractor states’ (124). Thus, cancers are often described as caricatures of their normal tissue architecture, with phenotypic states that may be less stable and more susceptible to heterogeneity in the TME than their normal tissue equivalents (124).

As cancers develop, tight regulation of the ECM is lost and tissue architecture begins to degrade (125). Solid state ECM interactions are necessary for cells to maintain potency and regulated ECM helps maintain the stem cell niche (126). As cancer is often associated with a blurring of the boundaries between stem and differentiated cells, it is possible that a loss of structured ECM is essential for the stability of multiple sub-dominant cellular clones within a tumour (124). A recent study by *Wang et al.*, provides direct evidence that ECM dependent signalling confers dynamic switching between transforming growth factor β receptor 3 (TGFB3) and jun D proto-oncogene (JUND) related expression signatures (127). Hence, depending on ECM stiffness and composition, heterogeneity in signalling pathways could be generated across a tumour.

More broadly, post-transcriptional regulation of gene expression, for example microRNA modulation of mRNA-mRNA interactions, can have profound consequences on context-dependent heterogeneity, for example in cell adhesion and Wnt signalling¹⁰ (99). Moreover, cells may be able to enter transient drug-tolerant states based on, for example, histone demethylase-mediated chromatin remodelling¹¹ (86). In summary, cells exist not in isolation, but in a milieu growth factors, cytokines, morphogens, and biophysical cues. Through contact guidance, differential ECM features are able to drive the initial complexity of a developing organ, and could have similarly profound effects on the context specific regulation of a tumours biology (126).

¹⁰ Two of the behaviours described by *Dvinge et al.*, by profiling the expression of 850 miRNAs in the METABRIC cohort (99).

¹¹ As seen in reversible Jarid1A-mediated erlotinib resistance in PC9 cells (86).

1.4 Clinical Implications of Tumour Heterogeneity

The typical attrition rate of new investigational drugs submitted for clinical trials is around 88% (128). Consequently, the average cost of bringing a new therapeutic agent through to regulatory approval, a process that can take a decade, is over \$2.56B (128,129). In order to reduce the attrition of experimental cancer agents and improve the outcome of patients treated with approved targeted agents, we must develop a more comprehensive picture of, and pre-clinical models able to capture, tumour heterogeneity.

1.4.1 Stratified Medicine

The first and most profound consequence of tumour heterogeneity for clinical practice is that chemotherapy and targeted agents do not have uniform efficacy across malignancies of the same subtype or even across the same tumour (130). In breast cancer, for example, the earliest subtype stratifications were defined by the presence or absence of hormone receptors (ER α /PR) and HER2. This first molecular stratification had unprecedented clinical implications, exemplified by the strong benefit of oestrogen pathway inhibitors in ER α +, and anti-HER2 therapy in HER2+, breast cancers. With the advent of large scale sequencing projects, our stratification of breast cancer has become more precise (131). Early genomic classifications based on single parameter¹² have evolved into complex integrative methodologies designed to capture heterogeneity across multiple levels, such as the 11 Integrative Clusters defined by *Curtis et al.*, (7), see *Figure 1.4*. Multi-parameter stratification continuous to improve. Now, efforts are underway to stratify both breast (132) and colorectal cancers (133) based on immune infiltrates and immunogenomic signature. Such classification could allow more rational use of, for example, novel immunotherapies (43).

¹² For example 50-gene PAM50 expression signature used to define the four intrinsic breast cancer subtypes (luminal A, luminal B, basal-like and HER2-enriched) still used in many research studies (8,9).

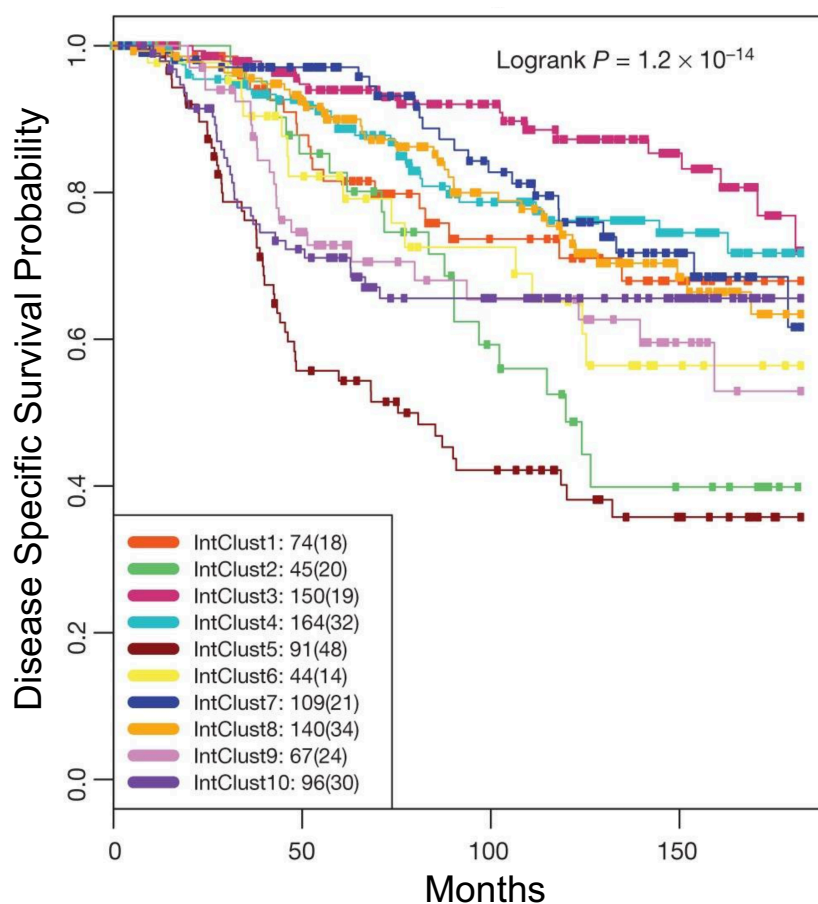
Figure 1.4 – Integrative Clusters defined in the METABRIC cohort**Figure 1.4** – Integrative Clusters defined in the METABRIC cohort.

Figure 1.4 shows a Kaplan–Meier plot of disease-specific survival (truncated at 15 years) for the ten original integrative cluster defined in the METABRIC cohort. For each cluster, the number of samples at risk is indicated as well as the total number of deaths (in parentheses). Figure is adapted from Curtis et al., (7).

1.4.2 Resistance to Cancer Therapy

Differences in clonal populations within the same tumour can have a profound influence on response to therapy, the emergence of drug resistance and disease progression. Currently our ability to predict the emergence of drug resistance in tumours requires *a priori* knowledge of resistance mechanisms and the identification of resistance-associated clones within a tumour. However, there is some evidence that the development of resistance is an inevitable consequence of single agent targeted therapies (134), suggesting a more universal strategy for identification of relapse risk may be possible.

Typically, resistance results from the outgrowth of specific pre-existing populations within a tumour rather than from *de novo* evolution (135). Indeed, the wider the diversity of minor clonal populations in a tumour, the more likely it is that resistance will arise. Such an association between tumour heterogeneity and drug resistance has been noted in ovarian (136), and oesophageal (137) cancers. Additionally, basal-like TNBCs have previously been linked with shorter disease free survival compared to non-basal-like TNBCs and tend to be associated with higher clonal diversity (78). Thus, there may be multiple routes to the development of resistance within any single tumour. Epidermal growth factor receptor (EGFR) is a well-established driver of CRC and anti-EGFR therapy shows clear benefit in a subset of the metastatic disease. However, a plethora of events have been shown to predict drug sensitivity (primary resistance) and acquired resistance to anti-EGFR therapy in this setting (134).

Interestingly, resistant populations have been shown with mutations in *RAS*, *BRAF* and *PIK3CA* or amplifications in *KRAS*, *ERBB2* and *MET*. While the mechanisms of resistance are genetically heterogeneous, they functionally converge on key signalling pathways which might aid the identification of biomarkers of disease progression (138). Similarly, numerous avenues to poly ADP ribose polymerase (PARP)-inhibitor resistance have been described in breast and ovarian cancers in either a BRCA1 dependent or

independent (*e.g.* 53BP1/REV7 loss) fashion (139,140). Each mechanism of resistance results in a clone regaining the ability to undergo homologous recombination, suggesting that functional biomarkers of resistance may be possible.

Genomic instability is the driving force of tumour heterogeneity. Although intratumour heterogeneity is linked with poor patient outcome, genomic instability is only associated with poor prognosis to a point. A recent study examined 1,000 treatment-naïve tumours and found that the total number of genomic clones had significant association with overall survival (141). However, the authors note that high clone number was only indicative of survival up to a maximum clonal diversity of four. Indeed, a diversity of more than four sub-clones was associated with longer overall survival (141). The authors used a 10% cell frequency cut off in their studies, yet, it is rare clonal populations which are thought to have evolved most recently (142) and may be more associated with resistance to targeted therapy¹³ (143–145). This could go some way to explaining the apparent discrepancy seen between this, and other studies.

Although anti-cancer therapies introduce a strong selective pressure, they do not necessarily lead to a reduction in overall clonal diversity or tumour genomic heterogeneity (92). For example, in a study of 47 breast cancer patients, strong changes in cellular phenotype were seen before and after chemotherapy, with no corresponding changes in genetic diversity, implying that a shift in the epigenomic landscape had resulted from exposure to chemotherapeutic selective pressures (146). In addition, several studies have identified the role of transient epigenetic states in the resistance to cancer therapy. For example, *Sharma et al.*, consistently detected a subpopulation of cells with >100-fold reduced erlotinib sensitivity across a panel of eight cancer cell lines (86). The authors found that this drug-tolerant phenotype was transiently acquired and lost

¹³ Conceptually, a clone may be at a fitness disadvantage relative to the tumour whole in the treatment naïve scenario, limiting clonal expansion. With the addition of temporal selective pressure (*i.e.* treatment) the clones relative fitness may increase, causing clonal expansion and eventual dominance.

by individual cells within the population in a process linked to IGF-1 signalling and histone demethylase-mediated chromatin remodelling (86).

Hence, both intra- and intertumour heterogeneity have profound clinical consequences in terms of differential response to therapy, development of drug resistance and disease progression. Beyond stratified medicine, a better understanding of the causes and consequences of clonal heterogeneity within a tumour will allow a deeper understanding of the emergence of drug resistance. New analysis tools such as the REVOLVER package could empower researchers to stratify patient groups based on the basis of how their tumour evolved (81) and perhaps allow prediction of a tumours evolutionary trajectory and a corresponding therapeutic strategy. Moreover, a greater understanding of genomic instability and its contribution to treatment resistance, and sensitivity, is needed.

1.4.3 Metastatic Progression

Metastasis is the ultimate cause of 90% of all cancer deaths (147). Aside from the development of resistance and recurrence of disease, heterogeneity among cancer cells widens the diversity available for the evolution of metastatic populations. The long-standing observation that some cells within a tumour were able to form secondary tumours at a higher frequency than others was one of the key arguments for the CSC hypothesis (17). However, multi-region sequencing studies have found that multiple distinct genomic clones are able to form metastases in pancreatic cancer (148), suggesting that a single 'CSC clone' is not necessary responsible for cancer dissemination.

Metastasis is thought to be the end result of a multistage process that includes local invasion by the primary tumour cells, intravasation into the blood or lymphatic system, survival in circulation (hematogenous and/or lymphatic), arrest at a distant organ,

extravasation, survival in a new environment, and metastatic colonization (125,149,150). Each of these steps relies on specific phenotypic features of the tumour cell, as well as interactions with the host microenvironment and the immune system (151,152). The long-standing observation that some cells within a tumour were able to form secondary tumours at a higher frequency than others was one of the key arguments for the CSC hypothesis (17), *i.e.* a cell able to possess a phenotype amenable to each stage in the metastatic process was likely possessed of some transcriptional plasticity. However, multi-region sequencing studies have found that multiple distinct genomic clones are able to form metastases in pancreatic cancer (148), suggesting that a single 'CSC clone' is not necessary responsible for cancer dissemination.

There are two general models of metastatic dissemination: the linear progression model and the parallel progression model (150). Both models assume that the primary tumour and its metastases are clonally related, in that they derive from a common ancestral cell. In the linear progression model, metastases emerge from late occurring advanced clonal subpopulations (153). The parallel model suggests that a metastasis is seeded early in molecular time from the primary site and progresses through the independent acquisition of mutations which may be different to the primary (124). A third mechanism, the cascade hypothesis, suggests that a polyclonal metastasis is in part due to direct metastasis-to-metastasis seeding (154,155).

Recently, *McPherson et al.*, performed phylogenetic analysis of 68 samples from seven patients with high-grade serous ovarian cancer (156). Through phylogenetic tree reconstruction from whole genome and single cell sequencing, the authors find that both unidirectional monoclonal seeding and polyclonal spread with reseeding contributed to the formation of intraperitoneal metastatic deposits. Importantly, only a minority of patients exhibited extensive migration and reseeding of polyclonal mixtures at multiple sites, implying that selection of clones in the peritoneal cavity was non-uniform. The

authors propose that this observation is linked to specific microenvironmental properties of the peritoneal cavity itself (156).

In another study, *Yates et al.*, investigated the patterns of genomic evolution between primary and metastatic breast cancers, across 299 samples from 170 patients (157). The authors found that genomic clones seeding metastasis primarily disseminated from late primary tumours, but continue to acquire mutations through similar mutational processes as their corresponding primary (157). These results potentially reconcile the linear and parallel models for metastatic spread (150). Indeed, most distant metastases acquired driver mutations, many of them clinically actionable, not seen in the primary tumour.

Hence, there is still much confusion regarding the presence of a metastatic clone or set of clones, with studies in ovarian cancer favouring monogenomic spread from primary tumours and in breast cancer polygenomic seeding, with further parallel evolution, from late stage primary tumours. Additionally, new research suggests that metastatic sites must be 'primed' before disseminating cells can form distant metastasis (158,159). It is possible that clonal cooperation could contribute to this effect, with one cellular population releasing cytokines and the other disseminating into the circulation. In summary, though much is still unknown, by studying the evolution of clonal populations we may be able to predict, and ultimately counter, the emergence metastatic clonal populations within a tumour.

1.5 Breast Cancer Tumour Initiating Cells (BC-TICs)

The CSC hypothesis was developed in part to explain intratumour heterogeneity. According to this hypothesis, many cancers have a unique subset of cells, referred to as CSCs or tumour initiating cells (TICs) that have the capacity to self-renew and give rise to other cancer cells, creating a hierarchically organised tumour (160). Over time, the theory has been adapted and developed to explain various properties of the tumour; CSCs have been linked to disease relapse, the development of drug resistance and the seeding of metastases.

In breast cancer, early fluorescent cytometry studies showed that a CD44⁺/CD24^{-/low} population could be isolated from bulk tumours and had enriched tumourigenic potential in limiting dilution transplantation assays (LDA) in immunocompromised murine models (161,162). This putative BC-TIC population maintained its tumour-initiating advantage over multiple passage and could repeatedly reconstitute phenotypically heterogeneous tumours (161,162). More recent work has implicated further surface markers (Protein C Receptor (PROCR), alpha-6 integrin (CD49f), Thy1⁺/CD24⁺ *etc.* (163–165)) in various breast cancer models. Cells with high aldehyde dehydrogenase (ALDH) activity have been shown to be enriched in BC-TIC content, forming xenografts in nude mice at 1000-fold efficiency of ALDH-negative populations (166). However, it is important to note that expression of these stem cell markers varies between breast cell lines and primary tumours and these markers do not universally enrich for BC-TICs (167).

Though LDA assays remain the gold standard for functional identification of BC-TICs, mammosphere forming assays (MFAs) have been developed and widely adopted as an *in vitro* surrogate of TIC content (168,169). Pleural effusions isolated from breast cancer patients were shown to be highly enriched in both sphere forming ability and xenopplantation potential in immunocompromised mice, but there was no correlation

with CD44⁺/CD24^{low/-} populations. Similarly, in primary human breast cancers, CD44⁺/CD24^{low/-} expression is not an independent predictor of survival (170). The same study found that in 4,125 patients CSC markers tended to associate with ER α - breast cancer and the BC-TIC markers ALDH1A1 (for ALDH positivity) and CD49f were of independent prognostic value for ER α - patients (171). In general, BC-TIC associated signatures are most prominent in TNBC, belonging to either basal-like or claudin^{low} subtypes and are indicative of a worse overall prognosis compared to luminal A tumours (159,172,173).

Perhaps most significant, however, is the association between BC-TICs and resistance to common therapies. Following chemotherapy, residual breast cancer cells are enriched in CD44^{high}/CD24^{low} subpopulations, have increased mammosphere forming efficacy (MFE) and increased tumour initiation potential by LDA (174–177). BC-TICs have been hypothesised to be quiescent in nature, demonstrating this, isolated BC-TICs show increased dye retention and a decrease in cell cycle related expression markers (178–181). This quiescent nature may confer on BC-TICs their demonstrated resistance to such drugs, since efficient induction of apoptosis by traditional cytotoxic chemotherapies requires cell division (175,182). Even in cell line models, CD44^{high}/CD24^{low} are slow cycling, enriched in MFE and are resistant to some chemotherapies. Additionally, *side-population*¹⁴ cells isolated from the MCF7 cell line are enriched for these BC-TIC features and for the expression of P-glycoprotein, a well characterised ATP-binding cassette (ABC) multidrug resistant transporter protein, potentially hinting at a more complex mechanism of drug resistance (181,183).

There is evidence to suggest that BC-TICs and non-TICs are inter-convertible, either spontaneously or through induction (184–186). This apparent plasticity in BC-TICs has also been hypothesised as an obstacle in treating breast cancer, although facilitating the transformation of TICs to non-TICs has been suggested as a novel therapeutic strategy

¹⁴ Defined as a distinct flow cytometry population with lower overall positivity for Hoechst 33342 staining.

(160). In breast cancer cell lines, a state of phenotypic equilibrium appears to exist, in that single CD44^{high}/CD24^{high} cells give rise to CD44^{high}/CD24^{low} progeny and *vice versa*, apparently following a stochastic pattern as predicted by Markov models (184,186).

Further demonstrating this plasticity, in transformed human mammary epithelial (HMLER) cells, the CD44^{low} subpopulation can convert spontaneously to a CD44^{high} phenotype *in vitro* and *in vivo*. The underlying mechanism of these interconversions has yet to be fully elucidated, but has been linked to the EMT transcriptional program and TGFβ signalling (187). For example, the *Zinc finger e-box1 (ZEB1)* promotor undergoes conformational changes in response to TGFβ signalling to drive breast cancer plasticity and the conversion of CD44⁺/CD24⁺ luminal-like cells to CD44⁺/CD24⁻ myoepithelial/basal-like cells is governed by Activin/Nodal initiated TGFβ signalling (186,188). Several caveats in these studies exist, for example EMT transcriptional programs have yet to be described in this context *in vivo* and TGFβ signalling is known to be highly context dependent in its action (159). For example, *Bruna et al.*, studied the effects of TGFβ on BC-TICs and EMT in a panel of breast cancer cell lines (159). The authors found that TGFβ increased MFE in claudin^{low} cell lines by orchestrating a specific gene signature enriched in stem cell processes and predictive of worse overall clinical outcome in breast cancer patients. Similarly, the authors found that TGFβ induced progenitor activity in normal mammary epithelial cells only in the basal/stem cell compartment, where claudin^{low} cancers are presumed to arise (159).

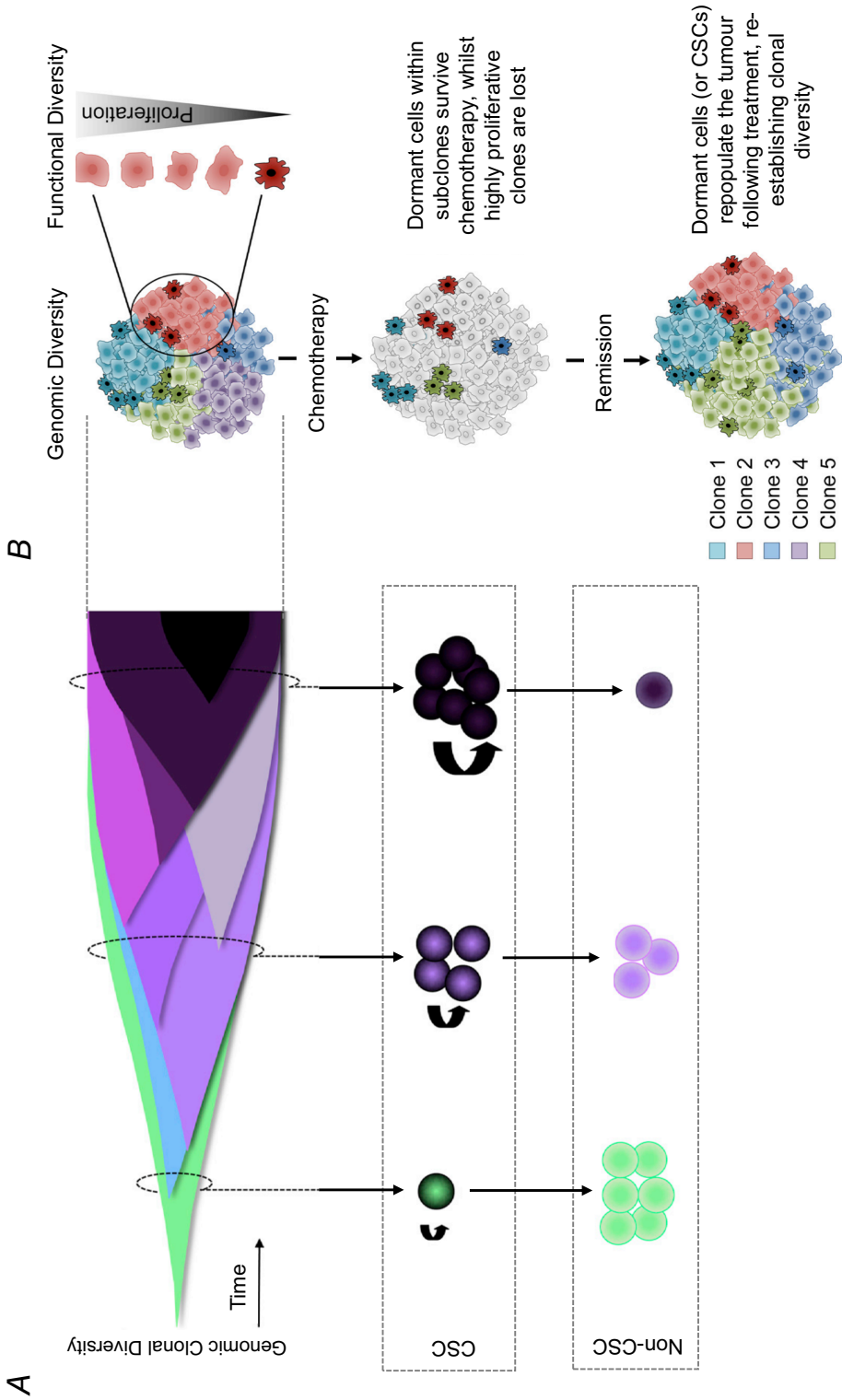
Recent research in the stem cell field has identified various tissues with dormant or quiescent stem cell populations with the ability to regenerate tissues on damage (189–191). These cells are resistant to normal genotoxic stress (of the type used in chemo- and radiotherapies) by virtue of their slow cycling nature. For these reasons and others, Kreso and Dick have recently put forth a unifying model of CSCs and clonal evolution of genomic clones (24). The authors propose that certain cellular phenotypes within genomic clones may adopt a ‘dormant’ state but be able to reacquire malignancy in a

context specific manner, *Figure 1.5*. Such cellular populations have been described in a CRC PDX models (83) where quiescent cellular clones were linked to the acquisition of chemoresistance and later found to express the putative stem cell marker polycomb group RING finger protein 4 (BMI1) (192).

The realisation that oncogenic driver mutations can lead to reacquisition of multipotency has profound implications for this model (71,72). Rather than resulting from oncogenic transformation of normal tissue stem cells (193), a CSC-like phenotype could be the result of specific mutational events in a particular genomic clone. Moreover, as cells may transition between different epigenetic attractor states in cancer, cellular dedifferentiation, reacquisition of multipotency and acquisition of a CSC-like phenotype could be driven by the epigenome (194). Thus, the CSC and clonal evolution models of tumour development can be reconciled by considering that genomic clones may contain a heterogeneous mixture of cellular phenotypes. These cellular phenotypes could be maintained by the diversity in epigenetic state and TME-mediated (spatially heterogeneous) signalling pathways across a tumour. This unified model reinforces the need to study tumour growth and heterogeneity in model systems that are able to retain the complexity of human malignancies.

Figure 1.5 – Kreso and Dick's Unified Theory of Tumour Heterogeneity. Adapted from (24). Figure 1.5 depicts the unified theory of tumour heterogeneity as proposed by Kreso & Dick (24). **A)** Top panel shows acquisition of favourable mutations resulting in clonal expansion of the founder cell. Bottom panel depicts CSCs as non-static entities able to evolve over the lifetime of a cancer. The figure also depicts CSCs as functionally different depending on genomic clone background. Specifically, developmental hierarchy may be steep (i.e. only few self-renewing CSCs exist among a large number of non-CSCs; left), or subclone may have high self-renewal potential, where most cells are tumorigenic. **B)** Clonal structures, depicted by different colours, contain a mixture of cells that differ in stemness and/or proliferative ability. Chemotherapy may reduce tumour burden by eliminating the highly proliferative cells within a subclone, as seen with Type IV clones in (24). **Figure Overleaf.**

Figure 1.5 – Kreso & Dick’s Unified Theory of Tumour Heterogeneity



1.6 Interrogating Tumour Heterogeneity

Multiple groups have attempted to define clonal dynamics based on either lentiviral tagging (cellular clones) or mutational clustering (genomic clones) by population and single cell based computational approaches¹⁵ (195–197). In our biobank of breast cancer PDX models, we rely heavily on PyClone for the inference of clonal structures during tumour repopulation and drug treatment (198). Additionally, seminal studies by both *Kreso et al.*, (colorectal cancer) (196) and *Nguyen et al.*, (breast cancer) (197) utilised lentiviral lineage tracing to mark the progeny of individual tumour cells in PDX models. Each group found a spectrum of clonal behaviours on serial xenograft passage and were able to draw clinically relevant conclusions from measures of tumour heterogeneity.

1.6.1 Genomic Clustering Techniques

Several computational methods exist to group somatic mutations based on shared frequency. These techniques attempt to reflect true genomically distinct populations within a tumour. For example, PyClone is a Bayesian clustering method for grouping sets of somatic mutations into putative clonal clusters. PyClone estimates cellular prevalence of clones and accounts for allelic imbalances introduced by segmental copy number changes (199). *Eirew et al.*, were able to reconstruct the genomic clonal dynamics of a panel of breast cancer PDX models using PyClone (195). In each of the 15 cases examined, clonal selection on engraftment was observed. This varied from extreme engraftment bias, selecting minor clones present in the sample of origin (<5% starting population), to only moderate clonal selection and polyclonal engraftment. Remarkably, similar clonal dynamics were observed in parallel xenografts established from the same

¹⁵ Strictly speaking, techniques allowing us to trace the progeny of single cells (such as lentiviral lineage tracing) are better equipped to define ‘clones’ than those designed to cluster somatic variations based on shared frequency. Even if such techniques were perfectly accurate, parallel evolution of the same genotype by two ‘clones’ in separate evolutionary trajectories cannot be ruled out. However, following convention, in our work, we will refer to genomic and cellular clones.

sample, suggesting deterministic kinetics of repopulation. In a separate study, *Ding et al.*, found PDX models established from a basal like breast cancer were more representative of the patient's metastatic lesion than primary tumour (153). *Eirew et al.*, further observed variable clonal dynamics between PDXs established from different molecular subtypes, underscoring the need for better representation of tumour molecular subtypes (7,195). In *Bruna et al.*, we study clonal architecture in individual breast cancer PDX samples and clonal dynamics upon engraftment and across serial passage in 104 samples from 22 distinct models (198). Consistent with *Eirew et al.*, we observe clonal selection on initial engraftment, but find only 20% of the distinct structures identified by PyClone change in cellular prevalence between passages. Interestingly, only 4 of the 38 clonal clusters that changed significantly after engraftment or during passaging contained a driver mutation¹⁶. These data suggest that proliferative advantage may not be central to clonal dynamics *in vivo* (198). Together, these observations suggest that deterministic, fitness-based mechanisms underline tumour-dependent clonal selection observed on engraftment.

1.6.2 Phenotypic Lineage Tracing Technologies

Multiple studies, across multiple areas of biology, have used the heritable nature of DNA to link mother and daughter cells in lineage tracing experiments. Typically, such technologies rely on viral DNA integration in the host cell, however detection techniques can be varied. For example, *Buczacki et al.*, used pulse-chase labelling of transgenic Cyp1a1-H2B-YFP mice to define quiescent label retaining cells in the intestinal crypt (189), whereas *Davis et al.*, used a multicolour lineage tracing approach, coupled with advanced 3D imaging, to uncover the unipotent nature of adult mammary stem cells (200). For longer term lineage tracing studies, researchers have typically relied on integration site analysis and/or high complexity nucleotide barcoding sequencing (201–

¹⁶ Driver mutations were defined based on *Vogelstein et al.*, (452). Genes in which these were found are: *ubiquitin carboxyl-terminal hydrolase (BAP1)* in STG139; *lysine (K)-specific demethylase 6A (KDM6A)* in HCI004; *mitogen-activated protein kinase kinase kinase 1 (MAP3K1)* in STG143 and *PIK3CA* in HCI008.

203). For example, *Nguyen et al.*, used high complexity barcodes to track individual cells in cell-line derived xenograft and PDX models of breast cancer (197). The authors focussed on interrogating the BC-TIC phenomenon, finding that ‘clone-initiating’ cells were present in the samples at frequencies of between $\sim 1/10$ and $\sim 1/10,000$ cells. Interestingly, in the cell-line xenograft models, clone-initiating frequency was negatively affected in transplants of more than 20,000 cells, suggesting that growth competition negatively affects xenograft diversity. The authors describe up to five clonal growth patterns when studying PDX repopulation dynamics (unchanging, expanding, diminishing, fluctuating or of delayed onset) (83,197).

Cell lines are often considered monoclonal, having lost diversity through numerous cycles of *in vitro* culture (204). *Porter et al.*, utilised cellular barcoding to simultaneously track the clonal dynamics in common cell lines including HeLa, K562 and HEK293-T (204). Each cell line exhibited ongoing clonal dynamics, even in optimal culturing conditions. Interestingly, the authors found that re-deriving K562 line from a single cell before barcoding and clonal tracking, reduced but did not eradicate clonal dynamics in the model (204). These experiments suggest rapid phenotypic-clonal dynamics are ongoing in commonly cultured cell line models.

1.6.3 Origins of Drug Resistance by Lentiviral Lineage Tracing

In two studies published in *Nature Medicine*, researchers used the ClonTracer barcode library to trace the origins of resistance to EGFR inhibitors erlotinib, WZ4002 and gefitinib (135,205). ClonTracer enables the labelling and tracking of >1 million individual cells with a unique 30-nucleotide long semi-random DNA sequence tag (135). The ClonTracer library was designed to have a balanced GC content to ensure uniform PCR-amplification efficiency and has a theoretical complexity of 73 million unique barcodes (135). Collectively, these studies show the power of lineage tracing using high complexity barcoding for interrogating the emergence of drug resistance.

In the first study, the library was used to track the emergence of erlotinib resistance in the non-small cell lung cancer (NSCLC) cell line HCC827 (135). HCC827 harbours an activating *EGFR* mutation that confers sensitivity to EGFR inhibitors and has been widely used to study resistance to this class of drug (206–208). The authors found, by next generation sequencing (NGS) of barcode amplicons, that 0.05% of HCC827 cells harboured a pre-existing resistance to erlotinib. Further analysis of this population found *MET* amplification, and treatment with the c-Met inhibitor crizotinib significantly reduced the diversity of resistant populations. A small proportion of cells were resistant to both inhibitors, and the authors noted that this sub-population tended to have a mesenchymal morphology. Gene expression profiling by RNA sequencing (RNA-seq) confirmed upregulation of EMT pathways, previously linked to EGFR resistance (209). This study went on to monitor the therapeutic response of the KCL-22 cell line to three ABL1 inhibitors. KCL-22 is a chronic myeloid leukaemia (CML) cell line derived from a patient in blast crisis and harbours the BCR–ABL1 translocation (210). The authors performed unsupervised hierarchical clustering on the 100 most enriched barcodes from each replicate across each treatment group and found clustering of barcodes enriched in nilotinib and imatinib treatment arms, but not GNF-2. These data suggest similar pre-existing clones drove resistance to imatinib and nilotinib (both catalytic subunit inhibitors), but that GNF-2 (an allosteric inhibitor) resistance was driven by separate cellular populations (211). Consistent with this notion, genomic analyses revealed that all of the imatinib- and nilotinib-treatment replicates harboured *T315I* mutations in *ABL1* on the population level, whereas all five GNF-2 replicates displayed an A337V variant in ABL1 (135).

In a second study, ClonTracer was used to further interrogate anti-EGFR resistance in NSCLC and delineate two separate evolutionary trajectories for resistance (205). Specifically, acquired resistance caused by the EGFR^{T790M} gatekeeper mutation can occur either by selection of pre-existing EGFR^{T790M}-positive clones or via genetic evolution of

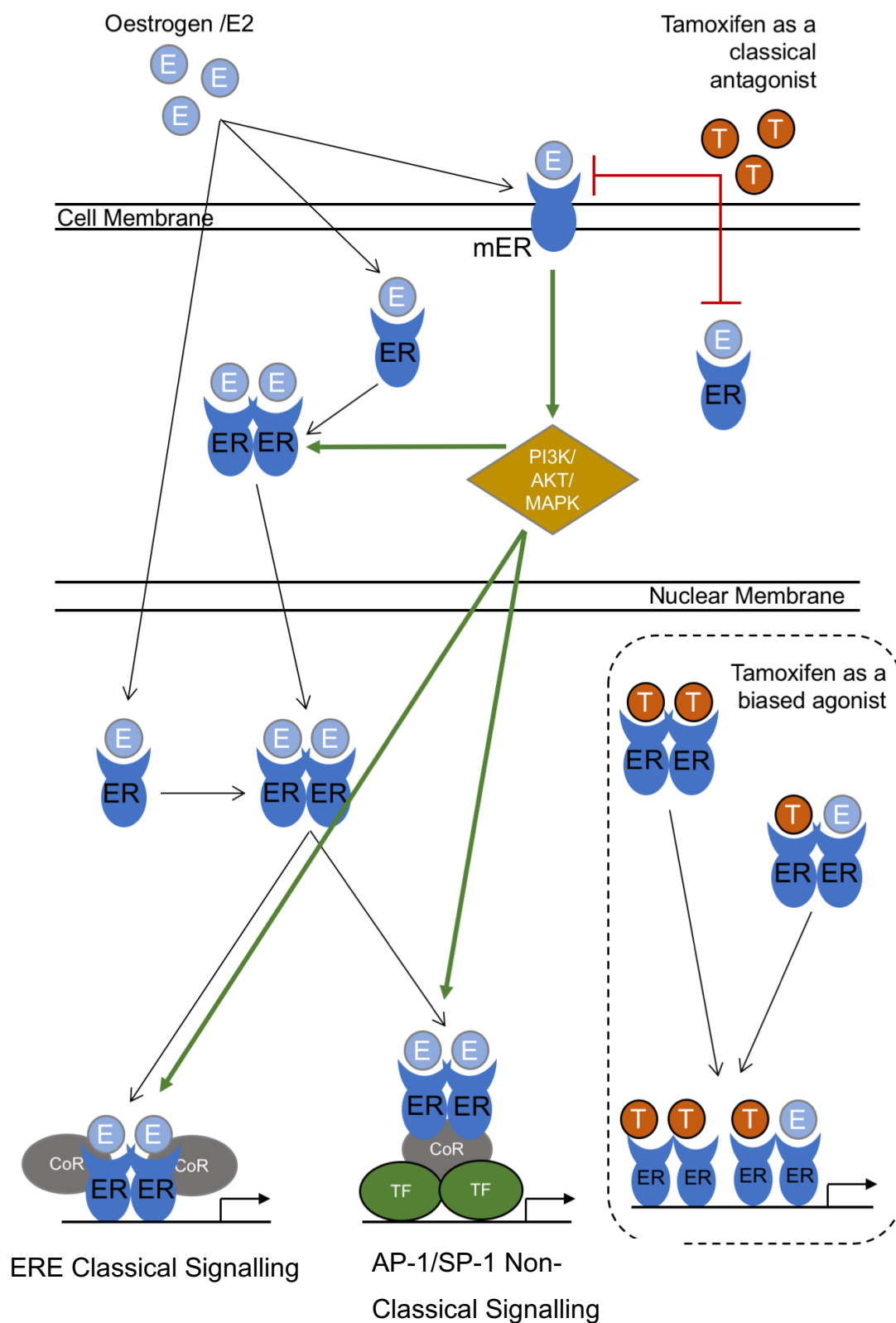
initially EGFR^{T790M}-negative drug-tolerant cells. Interestingly, these distinct paths to drug resistance were first hinted at by the variable time required to develop resistance in two PC9 sublines. On further investigation, the authors found that whilst both PC9-GR2 (6 weeks to resistance) and PC9-GR3 (24 weeks to resistance) had diminished upregulation of *Bcl-2-like protein 11 (BIM)*, a key mediator of apoptosis in *EGFR*-mutant NSCLC (211), induction of BIM protein levels after drug treatment was significantly lower in PC9-GR3 cells than in PC9-GR2 or parental cells. Additionally, WZ4002 elicited a cytotoxic response *in vitro* and induced tumour regression *in vivo*, in PC9-GR2 but not PC9-GR3 cells. Cellular barcoding studies using the ClonTracer library helped delineate that pre-existing resistant clones harbouring EGFR^{T790M} mutations were present in the bulk PC9 population, but that these clones were distinct from clones initially found to be EGFR^{T790M} negative by allele specific PCR, and later shown to develop EGFR^{T790M} mutations. Although initially the acquisition of a resistance mutation in the time scale of the performed experiments seems surprising, the authors mathematically modelled *de novo* acquisition of the EGFR^{T790M} mutation during drug treatment and calculated the fraction of drug-tolerant pools that were predicted to acquire mutations during 16 weeks of gefitinib treatment. Using a variable set of parameters for mutation and cell division rates, the model predicted emergence of EGFR^{T790M} mutations over a time period corresponding to several months, in line with results found experimentally (212).

1.7 Oestrogen Receptors in Breast Cancer

Oestrogen and its primary receptor, ER α , play a major role in the initiation and progression of breast cancer. Indeed, 70% of breast cancer patients express ER α , PR, and/or oestrogen-responsive and ER α -dependent gene products (213). ER α status was first recognised as an independent prognostic factor for early recurrence in breast cancer as early as 1977 and 40 years later ER α remains the cornerstone of breast cancer stratification and clinical management (214).

Figure 1.7 – Oestrogen and tamoxifen action at ER α .

Diagram shows the action of oestrogen (E) and/or estradiol (E2) action at ER α , together with classical models of tamoxifen (T) antagonism and possible biased agonism at the receptor. mER is membrane bound ER, CoR is coregulator and TF is transcription factor. Green arrows denote stimulation and red inhibition, grey arrows represent the flow of ligands and/or ligand bound receptor through the cytoplasm. **Figure Overleaf.**

Figure 1.7 - Oestrogen and tamoxifen action at ER α 

1.7.1 Oestrogen Receptor Structure and Function

Oestrogen's cellular function is mediated by two classes of receptor; ER α and ER β , each a member of the 48 member nuclear receptor (NR) superfamily (215) of ligand-regulated transcription factors. NR ligands, including oestrogen, progesterone, retinoic acid, oxysterols and thyroid hormones, are generally lipophilic and cross the plasma membrane to directly interact with NRs inside the cell (216). NR regulated transcription controls a variety of cellular processes including proliferation, development and metabolism. In addition, oestrogen has been shown to act through ER α in the cytoplasm of certain cell types to rapidly activate signalling pathways directly; such as in regulating vascular tone in endothelial cells (217).

Both classes of ER share several evolutionarily conserved functional domains, including the central DNA-binding domain (DBD), involved also in DNA recognition (218). Transcriptional activation is dependent on two distinct activation functions (AFs), AF-2 resides in the COOH-terminal ligand-binding domain and AF-1 (which is constitutively active) in the variable NH₂ terminus (219). On ligand activation, cell-specific transcriptional response to oestrogen is governed by a plethora of associated coregulatory proteins and pioneer factors, allowing context dependent effector functions of oestrogen signalling in a given cell or tissue (220). Recruited coregulators (CoRs) carry out all the reactions required for the entire transcriptional process; different receptors binding to the same genetic sequence can recruit different coactivators and thereby provide different transcriptomic response. Similarly, different ligands occupying the same receptor at the same site can induce different structural conformations in that receptor and lead to recruitment of different coactivators, and consequently, different gene expression patterns (221).

More than 450 coregulators of steroid hormone NRs have been reported in the literature (222). The Src family are the best characterised ER α coactivators. On primary recruitment by hormone bound ER α , Srcs serve as bridging molecules to bring in

coregulators such as p300 (a histone acetyltransferase) and Coactivator Associated Arginine Methyltransferase 1 (CARM1) (a histone methyltransferase) (223,224). Src-1 and Src-3 also contain intrinsic acetyltransferase activity toward histones (225). Interestingly, genome-wide chromatin Immunoprecipitation followed by sequencing (ChIP-seq) experiments in MCF7 cells have shown that genes regulated by Src-3 are not shared with the other two Src proteins, suggesting that they play non-redundant roles in breast cancer maintenance (226). Indeed, Src-3 is amplified in 10% of breast cancers and forced expression in the mouse mammary gland causes tumour development (227).

In addition to coactivators, there are a group of transcription factors termed corepressors which oppose the action of coactivators in NR-mediated transcriptional regulation. Nuclear receptor corepressor 1 (N-CoR1) and nuclear receptor corepressor 2 (N-CoR2) are the best characterised members of this group (228,229). Each has no intrinsic enzymatic activity and instead acts as a scaffolding protein for recruitment of histone deacetylases, including histone deacetylase 3 (HDAC3) (230).

1.7.2 ER α Signalling in Cell Fate Decisions

The adult mammary epithelium comprises two predominant lineages; luminal cells forming the ductal system and milk-secreting cells of the alveoli and myoepithelial (basal) cells lining the ducts and helping in milk secretion during lactation (231). There is a great deal of debate currently surrounding the maintenance of each lineage in the adult gland. Whilst transplantation assays report bipotent progenitors in each compartment with complete regenerative potential, lineage tracing studies predominantly identify unipotent progenitors as responsible for maintenance of the gland in the absence of perturbation (200,231–233). It has been proposed that mammary repopulating cells (identified by transplantation) are distinct from mammary stem cells under physiological conditions. Functionally, it seems likely that, as identified in other tissues, there are a population of normally lineage restricted progenitors

capable of repopulating the entire gland in injury. Indeed, recent studies have demonstrated lineage restricted cell populations can regain multipotency by the activation of oncogenic PI3K signalling, highlighting further potential sources of complexity (71).

Independent of exact identity, it has been demonstrated that mammary stem cells are dependent on oestrogen signalling for self-renewal and maintenance. Indeed, ovariectomised mice show markedly depleted mammary stem cell number and repopulation potential *in vivo* (234). Whilst cells of the myoepithelial compartment are predominantly ER α -, those of the luminal epithelium may be either ER α + or ER α -. Interestingly, the majority of luminal progenitors and milk secretory cells are ER α -, suggesting a role for oestrogen mediated paracrine signalling between cells of each compartment (235).

In general, organogenesis, development and maintenance by stem cell populations is controlled by a network of interactions between key transcription factors and their transcriptional programs. ER α has been shown to play a central role within this network in the context of mammary gland development and maintenance, in concert with a group of pioneer factors: transcription factors that can directly bind to condensed chromatin and recruit other transcription factors and histone modifying enzymes (236,237). These pioneer factors can bring positive or negative effects on transcriptional programs depending on the specific cellular context.

Two ER α pioneer factors; GATA-binding protein 3 (GATA3) and Forkhead box A1 (FOXA1) are thought to play an essential role in mammary gland fate determination and in ER α cancer specific biology. FOXA1 is a member of the forkhead transcription factor family, found to bind at more than 50% of ER α binding sites (238). FOXA1 is highly expressed in luminal cells of the terminal end bud (TEB) where it plays a critical role in ER α function; *Foxa1* null mice are associated with a loss of TEB formation but no defects

in embryonic development (239). FOXA1 is thought to contribute to ER α functionality in mammary cell fate primarily by modulating basal expression and functional activity of ER α . Loss of *Gata3* similarly results in severe defects in mammary development due to failure of TEB formation. Furthermore, GATA3 has been shown to participate in luminal epithelial differentiation required for lobuloalveolar development and forced expression of GATA3 in mammary stem cell enriched populations promotes differentiation into luminal cells (234,237,239). Further, ER α induces GATA3 expression in luminal cells, implying there is an interdependence of FOXA1, ER α and GATA3 in the maintenance of luminal cells (239).

1.7.3 ER α as a Target of Cancer Therapies

The first targeted antiestrogenic therapy for breast cancer, tamoxifen, has been a cornerstone of the clinical management of ER α + breast cancers for the past three decades and is thought to have saved thousands of lives. Tamoxifen is a triphenylethylene derivative, selective ER α modulator (SERM) with significant clinical utility in hormone sensitive breast cancers (240). Tamoxifen itself is a prodrug, having relatively little affinity for its target protein. It is metabolised in the liver by cytochrome P450 isoforms CYP2D6 and CYP3A4 into the active metabolites 4-hydroxytamoxifen (4-OHT; afimoxifene) and N-desmethyl-4-hydroxytamoxifen (endoxifen), each of which has 30-100 times more affinity for ER α than the parent compound (241). The major clinical limitation of tamoxifen therapy is the development of endocrine resistance in a large proportion of patients, including almost all with advanced metastatic disease (238,242).

Tamoxifen, and later derivatives such as raloxifene act by competing for oestrogen binding at ER α . An alternative strategy is to reduce expression of ER α directly on breast cancer cells. The first selective oestrogen receptor degrader (SERD), fulvestrant, was approved by the FDA in 2002 as a second line therapy for those who have progressed following endocrine therapy (243,244). However, a more direct mechanism may be to

inhibit production of oestrogen itself; via ovariectomy in premenopausal women and use of aromatase inhibitors (AIs) in postmenopausal women. AIs work by inhibiting the action of aromatase, which converts androgens into oestrogens in a process called aromatization (245). In the adjuvant setting, AIs have proven more effective in prolonging disease-free survival and are quickly replacing the other endocrine therapies in the management of ER α + breast cancer. Ten-year analysis of the ATAC clinical trial published in 2010 confirmed superior efficacy and safety of anastrozole (a third generation AI) over tamoxifen as initial adjuvant therapy for postmenopausal women with hormone-sensitive early breast cancer (246).

1.7.4 Mechanisms of Resistance to Endocrine Therapies

Almost 50% of patients with advanced breast cancer develop resistance to tamoxifen. Although AIs have a slight clinical benefit over tamoxifen, acquired and *de novo* resistance is still common (245). Fulvestrant has demonstrated efficacy in patients who have relapsed for a second-time following treatment with tamoxifen or AIs, but further randomised clinical trials are needed (243,244,247). Despite their great successes over the past three decades, the major cause of lack of efficacy of endocrine therapies remains the development of resistance (213,238,248). Several mechanisms for acquired (*de novo*) and innate resistance to endocrine therapies have been proposed.

ER α expression is the main predictor of response to endocrine therapy, and lack of expression is the principle mechanism for innate resistance to hormonal therapy (249). Interestingly, several studies point to the possibility of reactivating silenced ER α expression and thus increasing sensitivity to endocrine therapies (250–252). For example, the co-treatment with inhibitors of DNA methyltransferase-1 (DNMT-1), such as 5-aza-2-deoxycytidine (AZA), and histone deacetylase (HDAC), such as Trichostatin A (TSA) and suberoylanilide hydroxamic acid (SAHA), induce ER α gene expression in ER α negative (ER α -) breast cancer cells and restore sensitivity to anti-oestrogens (253,254).

However, around a quarter of patients with acquired resistance to tamoxifen have reduced expression of ER α , but will respond to second line ER α inhibitors such as AIs or fulvestrant (255). This suggests that loss of ER α expression and function may not be the only mechanism of acquired resistance to tamoxifen.

Cross-talk between mitogenic growth factor and ER α -signalling pathways has been shown to play a major role in acquired resistance to endocrine therapies (256). For example, reporter gene constructs in tamoxifen resistant cell lines have indicated that the EGFR/ mitogen activated protein kinase (MAPK) pathway can result in phosphorylation of ER α AF-1 at serines 118 and 167 (257). Phosphorylation of AF-1 can lead to ligand independent ER α activation, coregulator recruitment and oestrogen regulated gene transcription even in the presence of tamoxifen-bound-ER α (258–261). Similarly, in MCF-7/HER2-18 cells manipulated to express HER2 alongside ER α , tamoxifen behaves as a full ER α agonist and stimulates cell growth through oestrogen regulated gene transcription. In this model, gefetinib, a selective inhibitor of EGFR, restores sensitivity to tamoxifen (262). Additionally, the HER2 tyrosine kinase inhibitor AG1478 has efficacy in tamoxifen resistant MCF7s and the anti-HER2 antibody trastuzumab improves survival in patients with ER α +/HER2+ breast cancer when administered alongside AIs (263). However, in this case it is unclear whether efficacy is due to inhibiting two independent cell growth pathways.

ER α can also associate with the Insulin like growth factor receptor 1 (IGF-1) receptor and p85 regulatory subunit of PI3K at the plasma membrane, resulting in Akt activation. Akt can then phosphorylate nuclear ER α at serine 167 (AF-1 domain) and drive ligand independent activation (264,265). Although a similar mechanism to ER α /HER2/EGFR, this observation could have profound clinical utility. Mutations in PIK3CA are the most common genetic abnormality identified in ER α + breast cancers (73). Indeed we have recently found an association between *PIK3CA* mutation and reduced survival in three distinct subtypes of ER α + breast cancer (73). Combination therapy with tamoxifen and

BEZ235 (a dual PI3K and mTOR inhibitor) enhanced apoptosis in four breast cancer cell line models, relative to either agent acting alone (265). In patients with advanced breast cancer who have relapsed on AIs, everolimus (mammalian target or rapamycin (mTOR) inhibitor) has some clinical benefit in terms of time to progression and disease-free survival (266).

Tamoxifen has complex actions on the ER α . In the breast, it acts primarily as an antagonist or selective agonist, but in other tissues such as the uterus, cardiovascular system and bone, it acts as a full ER α agonist (267). This nuanced activity is attributed to the various coregulatory proteins controlling ER α tissue specificity and transcriptional programs. The ER α coactivator Src-3 is amplified in 10% of breast cancers and is associated with poor overall survival (227,268,269). Tamoxifen has been shown to induce ER α -Src-3 interactions in HER2+ breast cancers. In this scenario, tamoxifen acts as a full ER α agonist and drives molecular crosstalk with the HER2 pathway (262). Tamoxifen also significantly increases the expression of Src-1 and Src-3, both of which are associated with the development of resistance (270,271). In addition, tamoxifen treatment increases co-localisation between Srcs and ER α in resistant cell lines and knock-down of either coactivator in tamoxifen resistant cell lines restores endocrine sensitivity (270,271).

By genome-wide ChIP-Seq, the Carroll lab has made major advances in characterising ER α binding profiles in tamoxifen resistance. Beginning in 2008, Hurtado and colleagues showed that both oestrogen-ER α and tamoxifen-ER α complexes bound directly to a *cis* regulatory region in the *ERBB2* gene (272). When ER α recruited paired box gene 2 (Pax2), *ERBB2* expression was repressed but on the loss of *PAX2*, Src-3 could turn expression back on and led to the development of HER2 mediated tamoxifen resistance. In a separate study, Hurtado and colleagues went on to map binding of the FOXA1 pioneer factor and ER α in tamoxifen sensitive and resistant models (238). The authors demonstrated that FOXA1 was required for ligand independent ER α -chromatin binding

in tamoxifen resistance and that FOXA1 was negatively regulated upstream by 11-zinc finger protein (CTCF). By studying ER α -binding events in both good and bad outcome ER α + patients, Ross-Innes and colleagues were able to reaffirm that tamoxifen resistant cancers were still able to recruit ER α to the chromatin (248). They found that those patients who would go on to develop tamoxifen resistance had a unique set of ER α -binding regions and that the acquisition of these binding regions was predictive of relapse. This study underscored ER α -binding as a dynamic process regulated by pioneer factors such as FOXA1. The authors suggest that development of tamoxifen resistance was not due to selection of a rare sub-clone but the FOXA1 mediated reprogramming of ER α -binding on a rapid timescale. Finally, in 2015 Mohammed and colleagues demonstrated that PR is capable of modulating ER α behaviour and antagonising the proliferative effects of oestrogen in ER α + cell lines (273). The authors found that progesterone treatment synergised with the effects of anti-oestrogen treatment in cell line and PDX models. This significant body of work demonstrates further that ER α chromatin binding is a complex process heavily regulated by coregulators and pioneer factors and that these proteins can mediate the development of drug resistance.

The central role of FOXA1 in the development of tamoxifen resistance was reconfirmed in a study by *Patten et al.*, (376). Through genome wide ChIP-seq in 47 metastatic and primary breast cancer samples, the authors showed a phenotypic clonal population with differential FOXA1 binding was responsible for the development of resistance to endocrine therapies (376).

1.7.5 Pluripotency Transcription Factors in Tamoxifen Resistance

The CSC hypothesis, as discussed earlier in this Chapter, posits that a small population of cells within a tumour are primarily responsible for seeding metastasis and the development of drug resistance (24,274,275). In breast cancer, CSC markers can be as controversial as they are in adult breast stem cells, which may contribute to conflicting

reports of their importance in this disease (167). Irrespective of the contribution of BC-TICs to normal breast cancer biology, there at least appears to be a link between the development of tamoxifen resistance and transcriptional programs typically deployed in CSCs and/or other stem cell populations.

Perhaps unsurprisingly given the role of ER α in normal breast development and stem cell differentiation, exogenous oestrogen has been shown to reduce the proportion of stem cells in the normal human mammary gland, and in breast cancer cells. *Simões et al.*, enriched for stem cells from reduction mammoplasties by *ex vivo* culture as organoids (276). The authors found that expression of three canonical master stem cell transcription factors, homeobox transcription factor *Nanog-delta 48* (NANOG), *octamer-binding transcription factor 4* (OCT4) and *sex determining region Y box 2* (SOX2), was higher in organoids than in differentiated cells from the same donor. Moreover, expression was reduced to a level more associated with differentiated cells by the addition of exogenous oestrogen. In MCF7 mammospheres, the authors found that overexpression of *NANOG*, *POU5F1* or *SOX2* reduced *ESR1* expression, the number of putative BC-TICs and their capacity for invasion. Moreover, tamoxifen treatment in this model increased the number of mammospheres formed, and this corresponded to increased *NANOG*, *POU5F1* or *SOX2* expression (276).

In a later study by the same group, it was found that tamoxifen resistant cells were enriched for mammospheres forming cells and showed higher expression of *SOX2*. The authors found that silencing of *SOX2* by siRNA reduced the size of the stem/progenitor cell population and restored sensitivity to tamoxifen (277). *SOX2* has a controversial role in several cancers. For example in two papers published in *Oncogene* three years apart (2014 & 2017), *SOX2* was first found to play an essential role in the self-renewal of melanoma tumour initiating cells and hedgehog-induced melanoma cell growth (278), before being shown by CRISPR-Cas9 to have no functional role in melanoma initiation, growth or metastasis formation (279). Authors of the second study posit that the

apparent contradiction could reflect a distinct stem cell program active in neural crest stem cells and during melanoma formation (279).

OCT4, encoded for by the *POU5F1* gene and another master transcription factor of embryonic stem cells required to reprogram differentiated cells into iPSCs (280), has been independently linked to poor prognosis in breast cancer (281). Specifically, *POU5F1* expression associates with the stem cell marker ALDH1 but not with markers of EMT. In survival analysis, *POU5F1* expression was independently associated with poor prognosis in 319 cases of invasive breast cancer and in the ER α +, but not ER α -, subgroup (281). Interestingly, *POU5F1* expression was particularly associated with poor clinical outcome in ER α + patients treated with tamoxifen. OCT4 has recently been described as a novel ER α -associated transcription factor involved in ER α recruitment to tamoxifen, but not oestrogen, associated transcriptional sites (282). Further experiments uncovered a mechanism by which Nkx3-1, the androgen regulated transcription factor predominantly localised in the prostate epithelium (283), basally repressed *POU5F1* expression in MCF7 cells. Interestingly, the authors went on to show by quantitative mass-spectrometry that tamoxifen treatment of MCF7s elevated Nkx3-1 degradation through a p38MAPK-dependent phosphorylation of the E3 ligase, Skp2 at serine-64 (282).

Interestingly, the third canonical transcription factor needed for derivation of iPSCs, *kruppel-like factor 4* (KLF4), is generally associated with a favourable outcome in breast cancer (280,284). *Jia et al.*, found that higher expression of *KLF4* correlated with increased tamoxifen sensitivity in patients, and was positively correlated with ER α activity. The authors went on to knockdown *KLF4* in MCF7 and BCAP37 cells, finding increased tamoxifen resistance. When ectopic expression was induced in T47D and TamR lines, the authors found suppressed growth, invasion and migration, together with increased responsiveness to tamoxifen. Mechanistically, KLF4 was shown to suppress ERK and p38 signalling, which were generally more activated in resistant lines, pointing

to a potential role for targeting KLF4/MAPK signalling in tamoxifen resistant ER α + breast cancer patients (284).

Recently, the DNA methylome of endocrine sensitivity was characterised by *Stone et al.*, (285). The authors found that in tamoxifen resistance, DNA hypermethylation occurred at oestrogen-responsive enhancers. This was associated with reduced ER α chromatin binding and consequently decreased gene expression of key oestrogen regulators. In another study by this group, promotor demethylation with 5-Azacytidine (5-Aza), coupled with E2, restored ER α -regulated gene expression in TamR lines and induced a significant anti-proliferative effect (97). Cumulatively, these results highlight a novel role for ER α response element methylation in the induction of tamoxifen resistance. Interestingly, embryonic transcription factors OCT4, SOX2 and KLF4 each have roles in regulating the methylome of stem and progenitor cells to control differentiation (286,287). Additionally, the oncogenic histone methyltransferase and key embryonic regulator (and functional enzymatic component of the Polycomb Repressive Complex 2 (PRC2)), EZH2, has been shown to contribute to tamoxifen resistance by silencing the expression of ER α cofactor *Growth regulating estrogen receptor binding 1* (GREB1) (288). Together, these results highlight a potential mechanism by which embryonic transcription factors contribute to regulation of ER α transcriptional activity and hence influence the development of tamoxifen resistance.

1.8 Pre-Clinical Models of Breast Cancer

As discussed in this Chapter, breast cancer is a collection of diseases with distinct biological traits and clinical outcomes (289). Thus, no individual model would be expected to completely recapitulate human breast cancer in its entirety. Nevertheless, multiple models of breast cancer have been established over the years; both patient-derived and artificially engineered. In this section, the models most often utilised in basic research are considered.

1.8.1 Breast Cancer Cell Lines

Breast cancer cell lines have found extensive use in the investigation of proliferation, apoptosis, migration and the BC-TIC phenomenon. The first breast cancer cell line capable of surviving in culture for longer than two months was isolated in Detroit in 1970 and named MCF7 (290). This ER α positive luminal cell line has been heavily relied on in the study of tamoxifen resistance, leading to predictive biomarkers of resistance in patients (248). Together MDA-MB-231 (a triple negative cell line), T47D (a luminal cell line) and MCF7 account for more than two thirds of all abstracts mentioning breast cancer cell lines (291). Experiments in cell lines were crucial in the development of one of the first targeted therapeutic agents launched in 1998; the anti-HER2 trastuzumab (Herceptin®), a humanized antibody that binds to the ectodomain of HER2 (292) and has demonstrated remarkable clinical impact in *ERBB2* amplified breast cancers. Cell lines have also helped to elucidate the mechanisms of primary and acquired resistance to trastuzumab and are still being used for a significant proportion of breast cancer research today. These early successes supported the use of cancer cell lines for both drug development and biomarker discovery (293).

The artificial 2D system of *in vitro* culture has many drawbacks and several attempts have been made to increase the relevance of these incredibly tractable models. A seminal paper published in 2003 by Al-Hajj and colleagues demonstrated the presence of TICs in

pleural effusions from breast cancer patients, which later were shown to be maintained in suspension as 3D spheroids called mammospheres (162,169). Accumulating evidence has supported the use of this system to better understand the biology of specific facets of breast cancer, drug resistance and metastasis (274,294,295). Mammosphere cultures have also been used to unravel molecular mechanisms of signalling networks, for example those underlying the apparently paradoxical role of TGF β in breast cancer (159). Using these 3D mammosphere cultures, our group has identified TGF β breast cancer subtype specific regulatory networks dictated by the epigenomic landscape (296).

In an attempt to combine the high trackability of breast cancer cell line models with the biological relevance of *in vivo* cultures, many have turned to cell line xenograft models. In ER α + disease, for example, human breast cancer cells injected into the mouse mammary stroma, alongside an implanted source of exogenous human oestrogen, has been adopted as a more biologically relevant readout of clinically relevant phenotypes (476). Recently, however, *Sflomos et al.* have shown that such cells respond to the high levels of TGF- β in their new microenvironment through basal reprogramming, perhaps explaining their relative insensitivity to estrogen¹⁷ (477). Interestingly, the same study also characterised a method of using intraductal injections to introduce ER α + breast cancer cells directly into the mouse mammary duct. *Sflomos et al.*, found this method was closer physiologically to the natural environment of human breast cancer cells (477). More broadly, the developed Mouse Mammary Intraductal (MIND) syngeneic models have since shown utility across a variety of human cancer subtypes, allowing researchers to model a host of clinical relevant phenotypes (478). As syngeneic models of human cancers can be used to study the immune response, they could be of particular utility as pre-clinical models for testing immune-oncology agents.

¹⁷ Implantation of estrogen pellets that produce 18–40 times the physiological levels of estrogen in mice.

1.8.2 Breast Cancer Mouse Models

Despite not always exhibiting typical histopathological phenotypes seen in human breast cancers, genetically engineered mouse (GEM) models have been used extensively to investigate tumour initiation and progression. GEM models generally fall into three distinct histopathological categories, those closely resembling non-GEM tumours, those with unique transgene-specific phenotypes and those that resemble human malignancies. The choice of gene promoter and the mechanism of induction greatly influence the histological phenotype of the resulting tumour and this needs to be taken into consideration for all GEM studies (297).

GEM models driven by the mouse mammary tumour virus (MMTV) promoter were used to characterize the effects of several now widely accepted oncogenes and tumour suppressors in breast cancer (including tumour suppressors *PTEN*, *BRCA1*, *TRP53* and oncogenes *ERBB2*, *MYC* and *CCND1*) (298). When combined with advanced intravital imaging, GEM models have also been used to elucidate the precise role of macrophages in breast cancer metastasis. For example, the Pollard lab has relied heavily on these models to show that the purported metastasis-associated macrophages (MAMs) are active promoters of the metastatic cascade rather than bystanders (299). Like syngeneic models of breast cancer, GEM models have the advantage of including native stromal compartments of a malignancy. However, a major limitation of early GEM models in particular is their tendency to form ER α negative tumours, a limitation that indicates the study of endocrine modulating therapies such as tamoxifen is more suitable in other platforms (300,301).

Heterogeneity within a clonally expanding tumour is thought to be maintained because of dynamic selective pressures in spatially distinct tumour compartments. However, a recent study has uncovered a network of inter-clonal cooperation maintaining intratumour heterogeneity in a Wnt driven MMTV GEM breast cancer model (82). The authors simulated targeted therapy by removing Wnt1 and found that relapsing basal

populations recruited heterologous Wnt-producing luminal cells to restore cooperation. Alternatively, tumours evolved to rescue Wnt pathway activation through some other mutational event. In each case, drug resistance occurred from the cancer cell autonomous compartment in a fashion that could not easily be predicted. If such inter-clonal cooperation exists in human breast cancer, this would underline the need for clonally diverse preclinical models. However, it is not clear how well such models represent human tumours; Wnt, for example, has a relatively minor role in most breast cancer seen in the clinic (7), despite heavy reliance on the MMTV GEM model in the study of breast cancer basic biology.

A major limitation of breast cancer mouse models has historically been their lack of representation of ER α ⁺ disease, with GEM models typically generating ER α ⁻ breast cancers (476,479). Due to the lack of sporadic and human-relevant cases of ER α ⁺ disease in mice, the earliest experimental models used the chemical carcinogen 7,12-Dimethylbenz(a)anthracene (DMBA) to induce carcinogenesis (480). Though used extensively in early research into ER α ⁺ cancers, rapid DMBA-induced mutagenesis made these models unsuitable to model the earliest driver events in human neoplasms.

p53 is a well-known tumour suppressor, often deactivated in human ER α ⁺ breast cancer. Consequently, one of the best-known GEM for ER α ⁺ breast cancer is the *Trp53* null mouse (481). Limitations with this model¹⁸, however, have led to the development of other GEM models. In 2012, for example, a *Stat1*^{-/-} mouse model was developed and characterised to reliable for ER α ⁺ breast tumours that were hormone dependent (482). Currently, this is arguably the best approximation of human ER α ⁺ breast cancer (476).

¹⁸ Notably the tendency for *Trp53*^{-/-} to form lymphomas in mice, necessitating the transplantation of *Trp53*^{-/-} cells in cleared mammary fat pads of *Trp53*^{+/+} mice, and the fact that *TP53* is only inactivated in ~30% of human ER α ⁺ breast cancers.

1.8.3 In vitro Patient Derived Models of Breast Cancer

Realizing the importance of the cancer cell autonomous compartment in driving therapeutic responses, many researchers have turned to organoid cultures to study a wide variety of processes in development and disease. Beginning in 2009 the Clevers lab showed that single leucine rich repeat containing G protein-coupled receptor 5 (LGR5) positive Intestinal Stem Cells (ISCs) could build crypt-villus structures *in vitro* without a supporting mesenchymal niche (302). Subsequent research by this lab has identified culture conditions for normal and malignant pancreatic (303) and liver (190) organoids, amongst other tissue types. Organoids are generally genomically stable over long term passage (304), though it is unclear whether mixed organoid cultures of primary tumours can truly recapitulate the complex clonal heterogeneity seen *in vivo*.

Recently, a biobank of 20 human CRC organoids was established and characterized by exome-sequencing, RNA expression analysis and high-throughput drug screening (305). The authors show CRC organoids largely recapitulate most features of the originating tumour sample, and the biobank captures most of the mutational and expression landscapes observed in large CRC studies. The authors screened these cultures using an 83 compound library to identify molecular signatures associated with drug responses (305).

1.8.4 Patient Derived Xenografts

Perhaps the model best reflecting the complexity of human malignancies is the patient derived xenograft (PDX) (23). In this model, breast cancer clinical samples are implanted and propagated in highly immunodeficient mice, typically NOD.Cg-*Prkdc^{scid} IL2rg^{tm1Wjl}/SzJ* (NSG) or NOD.Cg-*Rag1^{tm1Mom} IL2rg^{tm1Wjl}/SzJ* (NRG) strains. PDXs reflect the originating sample's morphological and molecular features (306,307). Genomically, they recapitulate most of the clonal architecture found in originating sample's and remain predominantly stable throughout serial passaging (195,198). The current

iteration of these models do lack patient-matched stromal compartments and by necessity, lack a functional immune system (23).

Early breast cancer PDX studies suffered from low transplantation efficiencies and consequently a limited diversity of models (298,308), for example one study reported only three ER α + models in a cohort of 32 stably transplantable PDXs (309). Clearly, to be useful as preclinical models the early bias towards aggressive TNBCs has to be overcome. In this context, a new protocol involving intra-ductal injection of cells in female mice has been developed with the hope to dramatically increase engraftment rates especially in less aggressive tumour samples (310).

Recently, a PDX-based drug screening program of unprecedented scale was reported. A large collection (n=1075) of molecularly annotated PDXs derived from the most common adult cancer types was shown to capture the genomic and transcriptomic features of tumours seen in the clinical population as a whole. The majority of PDXs in this collection were treated with a variety of targeted compounds in a strategy dubbed '1x1x1' for 'one animal per model per treatment'. This approach mimics the reality of human clinical trials, which do not allow for technical or biological replicates. One of the key findings of this study was that a population of PDXs mimicked the spectrum of human clinical responses, reinforcing the translatability of these models to predict population-based drug responses. Moreover, known mechanisms of resistance were identified by this strategy; for example three PDXs treated with encorafenib developed resistance through *BRAF* amplification, a clinically relevant resistance mechanism (311). It follows from these data that there is a strong rationale for performing drug screens in PDX models to investigate population-based treatment responses.

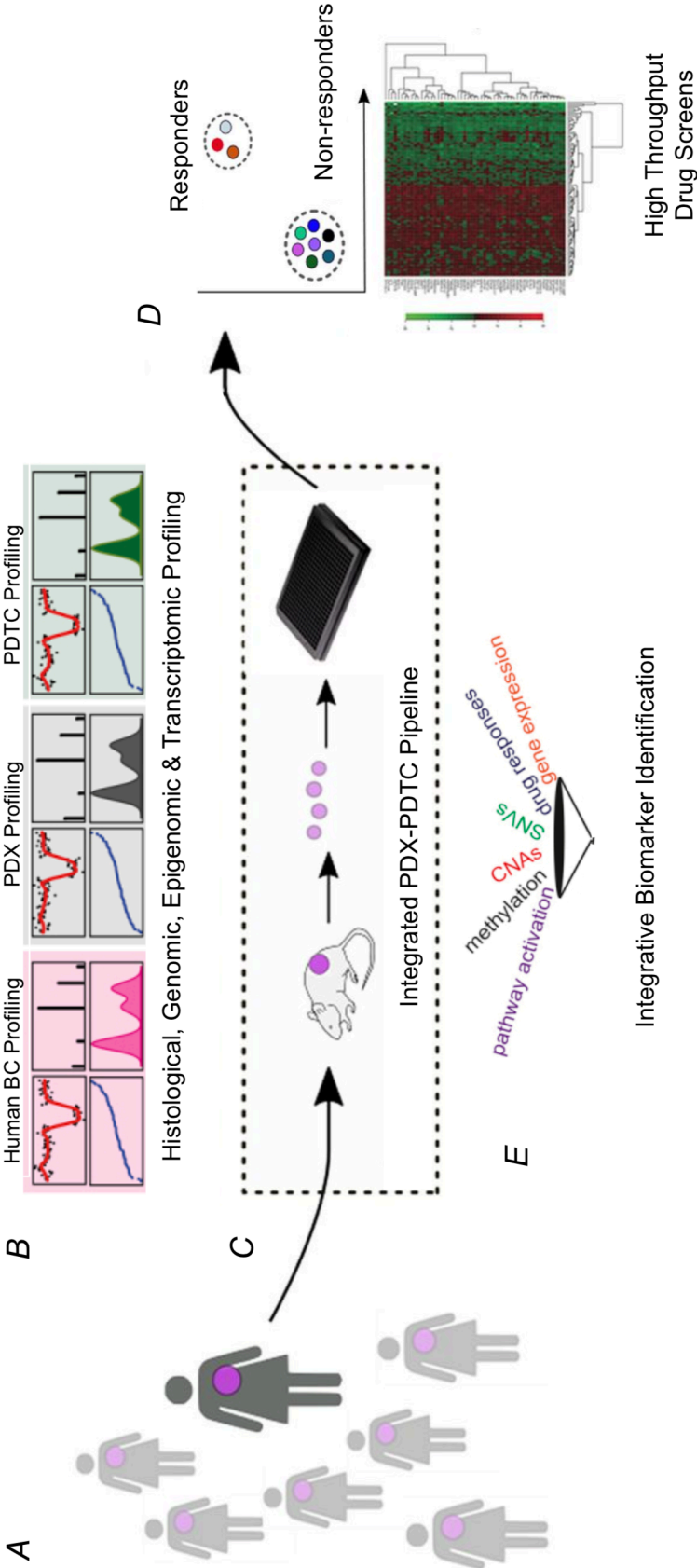
We have recently published a biobank of breast cancer PDX models with comprehensive molecular profiles and drug response profiles (198), *Figure 1.8*. Significantly, drug response data highly correlate across biological and technical replicates and compounds

with similar target specificities or mechanism of action cluster together. For example, similar responses were observed in 14 out of the 19 models tested with PARP inhibitor BMN-673 and Cisplatin (a DNA cross-linking agent), both of which exert their effects by increasing the frequency of mis-repaired double strand breaks in the absence of effective homologous recombination. Also, inhibitors of the PI3K-Akt-mTOR pathway shared a similar pattern of response across all samples tested. Recognising the importance of combination therapies in achieving long-lasting responses (312), we also tested and validated the PDX platform in a high throughput combinatorial drug screen with standard of care chemotherapeutic agents (Cisplatin and Paclitaxel) and six clinically relevant targeted compounds. Together this work validates the use of PDX models as an integral part of the breast cancer drug discovery and development pipeline. Crucially, as such models largely preserve heterogeneity found in their originating samples, they could be used to study the clonal drivers of resistance to therapy.

Figure 1.8 – Establishment of a large clinically annotated PDX cohort.

Figure 1.8 shows establishment of large clinically and molecularly annotated PDX cohort as described by Bruna and colleagues (198). Figure is adapted from (289) & (198). **A)** Depicts derivation of a highly diverse cohort of breast Cancer PDX models. **B)** All models were profiled for DNA, RNA, methylation and protein expression and heterogeneity found to be stable across breast cancer patients, PDX models and PDTcs. **C)** An innovative PDX-PDTC pipeline was developed for high throughput studies on PDX models. **D)** High throughput drug screens can be performed on PDTC biobank. **E)** Multiple analysis processes of control and treated PDX/PDTCs allow the identification of complexes, multi-omic, biomarkers of drug response or resistance. **Figure overleaf.**

Figure 1.8 – Establishment of a large clinically annotated PDX cohort



1.9 Summary

Diversity within a tumour has long been recognised. Indeed, as early as 1958, evolutionary biologist Julian Huxley commented on “genetic inhomogeneity” in cancer (313). It was not until the 1970’s that theories of tumour diversity were combined with Darwinian evolution to explore the causes and consequences of tumour evolution. Building on a study the previous year suggesting most cancers of the colon evolve through the now familiar polyp-cancer sequence (314), Peter Nowel put forward the first comprehensive theory describing tumour evolution in 1975 (12). Since this time, and particularly with the advent of high throughput NGS technologies, our understanding of the causes and consequences of tumour evolution has progressed at a rapid pace. Recent studies have uncovered the great diversity within tumours of the same organ, for example our own work in delineating the 11 distinct diseases typically classified as breast cancer (7,73), and the degree of heterogeneity seen within the same tumour (28,78). Tumour heterogeneity has profound consequences for the development of drug resistance and the seeding of metastatic sites, making its study of prime importance for the clinical translation of cancer research (11,130,315).

Crucially, tumour heterogeneity must be interrogated in our most widely used preclinical models of tumour biology. Recent studies have demonstrated the surprising complexity in even routinely cultured cell line models (135,204,205), but we must strive for the use of more complex models able to capture the tumour diversity of a human neoplasm. Recently, PDX models have been profiled extensively for clonal repopulation dynamics and engraftment biases (83,195,198). Our own work has shown that such models reflect most of the clonal structures of their originating samples and are amenable to a wide range of high throughput studies, including combination drug screens (198). However, even PDX models have major limitations, notably the lack of a functional human immune system, and so we must strive to view experimental results

within the context of the limitations of the model system used (23). Best practices and improvements to the PDX system put forward in (11) are outlined in *Figure 1.9*.

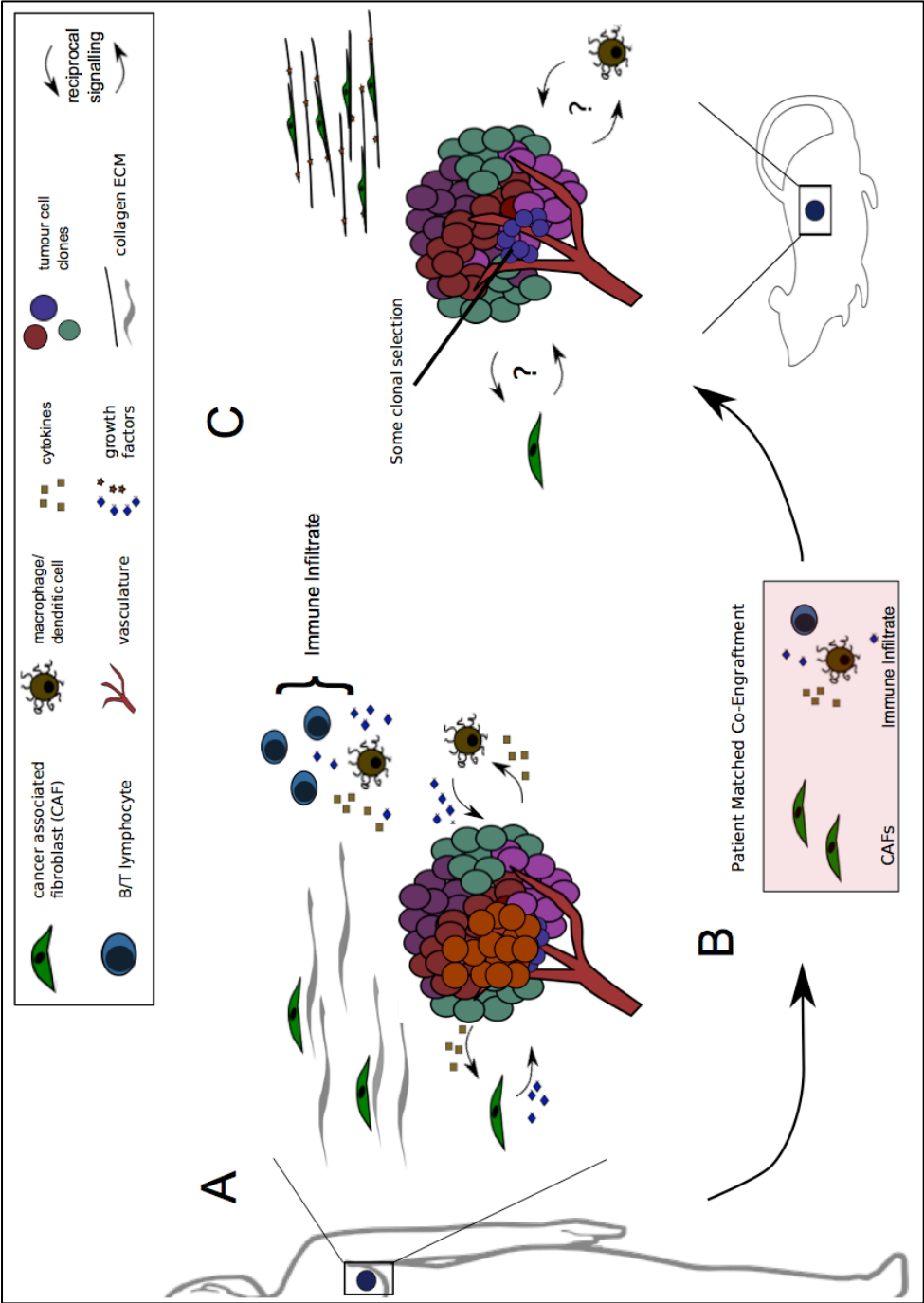
Breast cancer is the most common cancer diagnosed among US women (excluding skin cancers) and is the second leading cause of cancer death among women after lung cancer (316). Approximately 70-80% of breast cancers are ER α + tumours and are treated with oestrogen receptor antagonists (*e.g.* tamoxifen) or aromatase inhibitors (*e.g.* anastrozole). However, 50% of metastatic ER α + tumours will have innate resistance to such endocrine therapies and nearly 40% of initially responsive early-stage breast cancers will relapse with endocrine therapy resistant disease (317). Historically, tamoxifen was the most widely prescribed therapy for ER α + breast cancer, its continued clinical use means it remains as on the World Health Organisation's 2018 list of "Essential Medicines" (318). The development of tamoxifen resistance has been widely studied, but unlike many cases of resistance to targeted agents, its origins remain controversial (213,255,272,319,320). Crucially, the interplay between epigenetic (97) and transcriptional remodelling (238), embryonic and stem cell transcription factors (242,277,282,321) and resistance to tamoxifen therapy, promise to uncover a great deal about the biology and regulation of ER α in health and disease.

In this thesis, we describe experiments to uncover mechanisms of treatment resistance in breast cancer. First, we optimise a model to trace the clonal origins of drug resistance at single cell resolution, uncovering the isogenic but multiclonal origins of tamoxifen resistance in the MCF7 cell line. Secondly, we investigate several functional features of drug resistance models, uncovering a role for the embryonic transcription factor OCT4 in both the development of tamoxifen resistance and in the function of BC-TICs. Finally, we optimise a method of lentiviral lineage tracing in a PDX model, describing cellular clonal dynamics and laying the groundwork for drug resistance studies.

Figure 1.9 - Maintaining heterogeneity in PDX models.

A) Shows a primary breast tumour including some sources of heterogeneity found in the native microenvironment. Reciprocal signalling pathways between tumour cells and tumour associated macrophages (TAMs) and CAFs are highlighted. ECM is shown as collagen fibres with associated fibroblasts and macrophage/dendritic cells with T/B lymphocytes shown as part of the immune infiltrate. Other cell types (NK cells, myeloid-derived suppressors, etc.) have been omitted for simplicity. **B)** Patient-matched fibroblasts and immune cells as possible candidates for co-engraftment in the next generation of PDX models. Current PDXs established in NSG mice lack an adaptive immune system and may have impaired innate immune cell infiltrates and cytokine signalling due to defective IL2 receptor signalling (322). CAFs are known to contribute to treatment response, although murine fibroblasts are present in PDX models, it is unclear how faithfully these recapitulate their human counterparts (323). **C)** A PDX tumour in its native microenvironment. Questions over whether pro- and antitumor CAF/TAM signalling pathways are present to the same extent in PDX models as in the primary tumour are highlighted. Stromal and tissue architecture can have profound effects on transcriptional regulation but are often overlooked in the establishment of PDX models (126). To highlight potential differences in ECM organization between the native microenvironment and that of the PDX, here the ECM is shown as highly organized collagen fibres with closely associated myofibroblasts. Figure adapted from (23). **Figure overleaf.**

Figure 1.9- Maintaining heterogeneity in PDX models.



Chapter 2

Materials & Methods

2.1 Animal Studies

All experiments using animals were conducted under a UK home office license (Project License 707679) in accordance with the UK Animals (Scientific Procedures) Act 1986 and the European Union Directive EU 86/609 and underwent local ethical review at Cambridge University. Human tissues for PDX implantation were collected from consenting patients with appropriate approval by the National Research Ethics Service, Cambridgeshire 2 REC (REC reference number: 08/H0308/178). Some PDX models were obtained from our network of collaborators, with appropriate Institutional Review Board agreement and under Materials Transfer Agreements. NOD.Cg-*Prkdc^{scid} IL2rg^{tm1Wjl}/SzJ* (NSG) mice were obtained from Charles River. NSG, PDX and cell line xenograft models were maintained and passaged when necessary by the Cambridge Institute Biological Resources Unit (BRU).

2.1.1 PDX Biobank

Characterisation of our PDX models is covered extensively in *Bruna et al.*, (198). Briefly, a bank of human breast cancer explants has been maintained at the CRUK Cambridge Institute over the past six years, by combining efforts from Addenbrookes Hospital and collaborating hospitals in Europe (Institute Curie, Paris and VHIO, Barcelona), US (Huntsman Cancer Institute, Salt Lake City, Utah) and Canada (UBC, Vancouver). Surgically resected primary breast cancer tissue, biopsies from brain, skin, liver, bone, axilla and lymph node metastasis, and pleural effusions or ascites samples were obtained from consenting patients. Tissue samples were embedded in Matrigel and then implanted subcutaneously into two-four female severe immune compromised NSG mice. Pleural effusion and ascites samples were centrifuged, washed twice¹⁹ to eliminate red blood cells, and cell pellets resuspended in 50:50 Matrigel: Foetal Bovine Serum (FBS) solution before subcutaneous injection into mice (3-8 weeks old). Matrigel and FBS were batch-matched when possible in routine maintenance of the PDX biobank. Though it should be noted that the scale and long-term propagation of the biobank requires the use of more than one batch of each, which could introduce variability to our colony²⁰. It should also be noted that mice were not oestrous staged, which could introduce further variability to the colony. The time from patient collection to mouse implantation ranges from 30-180 minutes. PDXs were serially implanted into multiple hosts to allow *in vivo* expansion of each model. Xenograft samples were cryopreserved in liquid nitrogen and freezing media (FBS with 10% Dimethyl Sulfoxide (DMSO)) at each passage, from each mouse. Genotyping of all samples was performed to confirm matching with the originating patient derived sample. All models tested to date could be rescued by re-implantation of cryopreserved tissue. Clinical and molecular features of PDX models are shown in *Table 2.1.3*.

¹⁹ With Hank's Balanced Salt Solution (HBSS) with 155mM ammonium chloride, 0.1mM ethylenediaminetetraacetic acid (EDTA) and 10nM sodium bicarbonate, at pH 7.2.

²⁰ New batches of either Matrigel or FBS were batch tested by parallel growth of one PDX model in the previous and new batch, coupled with superficial inspection of heterogeneity and tumour growth curves to identify marked differences between batches.

PDX ID	Type	Grade	ER	HER2	PR	CK5	CK14	CK8	CK18	Ecad	Ki67	ESA Intensity	ESA %	pS6 Intensity	pS6 %	p53 Intensity	p53 %
STG143	Primary	2,3,3	+	0	-	-	-	+	+	+	low	3	30	3	NA	0	NA
STG195	Metastasis	3,3,3	+	2	+	-	-	+	+	+	high	3	70	3	40	3	30
STG201	Metastasis	3,3,3	+	0	-	-	-	+	+	+	high	3	80	2	10	3	100
STG282	Metastasis	3,3,3	-	0	-	+	+	+	+	+	high	3	100	3	90	3	100
STG335	Primary	3,3,3	+	0	-	-	-	+	+	+	high	3	100	2	10	3	100
HCI002	Primary	NA	+	0	-	+	-	+	+	+	high	2	20	2	10	2	90
HCI005	Metastasis	NA	+	0	+	-	-	+	+	+	high	1	5	3	80	2	30
HCI006	Metastasis	NA	+	0	+	-	-	+	+	+	low	1	10	3	60	2	10
HCI011	Metastasis	NA	+	0	-	-	-	+	+	+	high	3	100	3	80	0	0
VHIO098	Metastasis	NA	+	0	-	+	+	+	+	+	high	2	10	3	50	3	100
VHIO131	Metastasis	NA	+	0	+	-	-	+	+	+	high	3	100	2	20	3	70
VHIO244	Metastasis	NA	+	0	+	-	-	+	+	+	high	1	80	3	80	3	100
IC06	Primary	NA	+	0	+	+	+	+	+	+	low	1	80	3	20	3	40
IC06_TAMR	Primary	NA	+	0	+	+	-	+	+	+	low	0	0	3	20	3	80
IC07	Primary	NA	+	2	+	-	-	+	+	+	high	1	5	2	5	1	5

Table 2.1.3 - Clinical and molecular features of PDX models

Table 2.1.3 shows clinical and molecular features of PDX models utilised in this thesis. Assays are described fully in Bruna *et al.*, (198).

2.1.2 Generation of Patient Derived Tumour Cells (PDTCs)

Patient Derived Tumour Cells (PDTCs) were harvested from PDX tumours using dissociation mix²¹ by the following protocol. PDX tumours were harvested by dissection *post mortem* and minced into small fragments. Fragments were washed in HBSS with 5% BSA and 5% 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), incubated in dissociation mix at 5% CO₂ 37°C for 1-2hrs on a rotary shaker. Fragments were then washed as before and incubated in prewarmed Trypsin: Ethylenediaminetetraacetic acid (EDTA) (0.5% Sigma) for 5 minutes. Samples were washed and incubated with 2ml prewarmed 5mg/ml Dispase and 500µl of 1mg/ml DNase I. Samples were then washed and resuspended in 1:4 mixture of cold HBSS (5% BSA): ammonium chloride to encourage red blood cell lysis. Samples were again washed and filtered through a 40µm strainer. Viability and cell numbers could then be ascertained. PDTCs were maintained for up to 1 week in Mammary Epithelial Growth Media (MEGM) (Bulletkit; Lonza) in low adherent conditions (Corning ultra-low attachment plates) at 5% CO₂ 37°C.

2.1.3 In vivo Limiting Dilution Assay

Cells of interest were grown to 80% confluence before media was changed and ligand added if appropriate for 48hrs. 4-OHT at 300nM and E2 at 10nM. Following treatment, cells were harvested as normal by trypsinisation and resuspended 1 x10⁶ cells/ml. Viable cells were calculated by trypan blue exclusion and serially diluted, spun down and cell pellets resuspended in 50% Matrigel such that 50µl cell: Matrigel suspension would contain 1,000, 100 or 10 cells as appropriate. 50µl was injected into the 4th mammary fat pads of NSG mice along with 0.72mg/90day slow release/17B Estradiol pellets (Innovative Research of America) injected into the scruff. Injected regions were palpated

²¹ Dissociation mix: 2ml 10X Mouse collagenase/Hyaluronidase, 13ml Dulbecco Modified Eagle Medium (DMEM):F12, 5ml Bovine Serum Albumin (BSA) Fraction V (Gibco 7.5%), 10µl Insulin (final concentration 5µg/ml), 20µl Gentamycin (final concentration 50µg/ml)) (all Sigma Aldrich).

for signs of tumour growth twice weekly for up to 6months. We used the Extreme Limiting Dilution Analysis (ELDA) software package to interpret results (324).

2.1.4 Post-mortem Protocol for Identifying Metastatic Sites

At the end of the experiment, animals were euthanised in a CO₂ chamber and had their femoral artery severed, as per local procedures and best practises. In some cases, euthanised mice were inspected in the Xenogen small animal imaging system for suspected metastatic sites. Euthanised mice were placed in dorsal recumbency on a clean dissection board and an incision made across the full length of the ventrum (anus to chin), reflecting the skin and incising the abdominal wall, exposing the abdominal viscera, salivary and preputial/clitoral glands, and cervical and axillary lymph nodes. Skin was pinned back to expose the PDX tumour site (left or right flank). Tumour was identified, excised and stored as appropriate (neutral buffered formalin for histology, serum with 10% DMSO for cryogenic preservation and DMEM +10% FBS for further study, PDTC dissociation, dissection or reimplantation). Continuing with the *post mortem* examination, the rib cage was cut open to expose and examine the thoracic viscera by making two cuts laterally up each side of the ribcage, then one across, at the top of the sternum, to open a space wide enough to thoroughly examine all the lobes of the lung. All organs were visually examined for abnormalities. Specifically, the heart and lungs were identified in the thoracic cavity & liver, kidneys and spleen in the abdominal cavity. Any colour changes, size differences, and missing or mis-located organs were noted. The consistency of surfaces, any additional tissue (*e.g.* masses), fluid pockets, or the presence of fluid in the abdominal and thoracic cavities was noted. In some cases, organs with suspected metastatic deposits were further screened in the Xenogen small animal imaging system. We next checked the mesentery for enlarged lymph nodes and/or masses, harvesting the abdominal & inguinal mammary glands, together with suspected tumour draining lymph nodes (inguinal and/or lumbar).

2.2 Cell Culture Assays

T47Ds, MCF7s and HEK293-Ts were obtained from ATCC, TamRs were a kind gift from Dr J Carroll and were derived as described (325). Cell lines underwent regular screening for mycoplasma. MCF7s were maintained in DMEM: F12 (at 5% CO₂ 37°C) supplemented with 10% FBS²² and penicillin-streptomycin (PenStrep) (Sigma-Aldrich) and passaged by trypsinisation (0.25% trypsin EDTA; Life Technologies) every 7 days or at 80% confluence. TamRs were cultured under the same conditions, though with charcoal striped FBS (Life Technologies) and DMEM: F12 without phenol red to minimise oestrogens and pseudo-oestrogens. Prior to cell based assays, cells were briefly serum starved (media + 0.1% FBS) for 24hrs to synchronise cell cycles. Assay media containing 1% FBS was used in all 2D cellular assays. Viability was calculated by Trypan Blue exclusion assay; briefly, 0.1ml trypan blue (ThermoFisher) was added to 0.1ml cell suspension, the mixture was loaded into a hemacytometer and examined immediately under a microscope at low magnification (4x). Cell viability should be at least 95% for healthy log-phase cultures.

2.2.1 Mammosphere Growth Assays

Monolayer cells were trypsinised as normal and plated in Corning ultra-low attachment plates in mammosphere media (MM)²³. Mammospheres were cultured in low adherent conditions without phenol red at 1,000 cells/ml. Mammospheres were left to form over 7 days (1st generation spheres) with the addition of siRNA and/or treatment. Single cells were then generated by mechanical dissociation with the addition of prewarmed 0.5% trypsin: EDTA for 5-10 minutes. Cells generated from 1st generation spheres were

²² To minimise experimental variation, FBS, Charcoal-stripped FBS and BSA used in experiments were batch and lot matched.

²³ MM: DMEM/ F12 supplemented with BSA, 1x B27 (minus vitamin A), 1x insulin-transferrin-selenium, 0.5 mg/ml hydrocortisone, 2µg/ml human recombinant epidermal growth factor (EGF) (all from Life Technologies).

counted and scored for viability using a Vi-CELL instrument (Beckman Coulter; using manufacturers recommended protocols). Second generation spheres were then generated from 1,000 viable cells/ml. After 7 days culture in MM, wells were imaged using a Nikon Eclipse TS100 and mammospheres harvested for downstream analysis. 4x magnification images were taken to have some degree of overlap and stitched together in Adobe Photoshop CC 2014, ensuring no mammospheres were scored twice. Images were then exported to ImageJ64 (blinded by randomising file names) and mammospheres were manually fitted to circular outlines using the Region of Interest plugin (spheres are $>50\mu\text{m}^2$). Mammosphere forming efficiency (MFE) is calculated as number of 2nd generation spheres formed divided by number of viable cells plated. By capturing information from both MFE and mammosphere size, we hope to minimise the limitations of this assay – namely that mammospheres may be a product of multiple smaller spheres which aggregate together rather than the product of single cells. In some experiments we used a modified scale up protocol for mammosphere culture. Here, cells were trypsinised, singularised and seeded in 500ml volume of MM at 5×10^4 cells/ml density. The cell suspension was grown in 500ml spinner flasks with a rotating magnet within the flask for 7 days to generate mammospheres. This allows for mixing of the cultures constantly and prevents sticking of the mammospheres to the flask walls.

2.2.2 Wound Healing Assays

Single cells were plated in Essen ImageLock 24 or 96-well plates (Essen Biosciences) and cultured as normal to 90% confluence. Cells were then serum and oestrogen starved for 24hrs. Monolayers were washed with phosphate buffered saline (PBS) and scratched using a 24 or 96 well wound maker as per manufacturer's instructions (Essen Biosciences). If using 24-well wound maker, pipette tips were changed after each plate. Monolayers were again washed (gently in PBS) to remove dead cells and assay media (charcoal stripped, phenol red free 1% FBS, +/- treatment) was added. Scratched monolayers were cultured in an Essen IncuCyte FLR for up to three days with images

acquired every 2-4hrs. We aimed to avoid proliferation at the wound site to assay migration independent of proliferation. Thus, the assay was limited by visual inspection of dividing cells. Wound closure was recorded by the IncuCyte running the wound healing program. Relative wound density (measure of cells covering wound site) was used as the primary output of this assay.

2.2.3 Drug Screens and Dose Response Assays

Cells of interest were seeded in 24- or 96-well culture plates and grown to 90% confluency (~48hr). Ligand of interest was then added to experimental wells, and carrier to control wells. Ligands were: (Z)-4-Hydroxytamoxifen (4-OHT) ≥98% Z isomer (Sigma-Aldrich; 4-OHT) and β -Estradiol ≥98% (Sigma-Aldrich; E2). Tamoxifen is a prodrug that is metabolically activated by 4-hydroxylation to the potent primary metabolite 4-OHT or via another primary metabolite N-desmethyldtamoxifen (NDMTAM) to a biologically active secondary metabolite endoxifen through a cytochrome P450 2D6 variant system (CYP2D6) primarily expressed by the liver and certain areas of the central nervous system (*e.g.* substantia nigra) (326). Hence, direct use of 4-OHT over the tamoxifen prodrug was chosen. Plating was such that possible sources of variance due to well location was minimised by pseudo-randomisation. Where possible, outer wells were filled with PBS to further reduce variability. After the treatment period (typically 24-48hrs), media was aspirated, and cell viability quantified by CellTiter-Glo Luminescent Cell Viability Assay (Promega) as per manufacturer's instructions. Quantification of fluorescent signal intensity was performed using PHERAstar FSX Multimode Microplate Reader and inbuilt CellTiter-Glo assay protocol. In cases where multiple dose response curves were to be generated (for example in 4-OHT studies) a control (in this case parental MCF7) was included on each plate to identify inter-plate variation.

2.2.4 Proliferation Assays

Single cells were plated in Essen ImageLock 96-well plates (Essen Biosciences) at ultra-low density (2,000 cells per well) and allowed to incubate at ambient temperature for 20 minutes followed by 30 minutes at 37 °C in an Essen IncuCyte FLR system prior to scanning. Scanning and confluence calculations were carried out every 2hrs. After the first 4hrs culture, confluence was averaged across the two completed scans and used to balance ligand addition to wells. Wells with higher (>10%) or lower (<1%) confluence than normal were excluded from the experiment. Our aim was to ensure no ligand had an advantage by been added to high confluence only wells. Cells were incubated until confluence (up to 120hrs) and proliferation kinetics calculated in the IncuCyte software.

2.2.5 Matrigel 3D Growth Assays

Adherent cell lines or PDTCs were passaged as normal and resuspended at 1×10^4 cells/ml in appropriate media (supplemented with 0.1% FBS and 2% growth factor reduced (GFR) Matrigel, BD Biosciences). Cells were plated onto BD Falcon 8-well CultureSlides (BD Biosciences) pre-coated with 60µl collagen: Matrigel mix (collagen used was type IV from rat tail, Sigma) and incubated for up to 2 weeks at 37°C 5% CO₂, changing media every four days. Protocol was adapted from the Brugge lab at Harvard Medical School (327), available at <http://brugge.med.harvard.edu/>. Spheres were visualized by phase contrast microscopy or by 4',6-diamidino-2-phenylindole (DAPI) counterstain (DAPI dilactate, Invitrogen) and fluorescent microscopy. For 3D acinar growth assays, cells were trypsinised as normal and dissociated into single cells. 1×10^5 cells were resuspended in 75% Matrigel²⁴ (25% DMEM: F12) and plated as a 50µl droplet the centre of a 24 well plate. Wells where Matrigel touched the sides were disregarded. Matrigel was left to set for 2 hours at room temperature before MEGM was added. 3D acinar structures were left to form over 7 days before treatments were added. Images were collected with a Nikon Eclipse TS100.

²⁴ Matrigel used in these experiments was of the same batch and lot number.

2.2.6 In vitro Limiting Dilution Assays

Cells of interest were grown to 80% confluence before media was changed and ligand added if appropriate for 48hrs. 4-OHT at 300nM and E2 at 10nM. Following treatment, cells were harvested as normal by trypsinisation and resuspended 1×10^6 cells/ml. Viable cells were calculated by trypan blue exclusion and serially diluted to: Tube 1 (1,000 cells/well): 27,000 cells in 5.4ml media; Tube 2 (100 cells/well), 10,000 cells in 20ml media; Tube 3 (10 cells/well), 2,000 cells in 40ml media; Tube 4 (1 cell/well), 300 cells in 60ml media. Suspensions were thoroughly mixed before each dilution, before dispensing 200 μ l of each into sterile non-tissue culture 96 well U-bottom plates to give: 24 wells of 1,000 cells/well; 96 wells of 100 cells/well; 192 wells of 10 cells/well; 288 wells of 1 cells/well. Plates were topped up with growth media as appropriate and cultured for 3 weeks at 5% CO₂ 37°C. After 3 weeks, plates were scored for spheroid growth by microscopic visualisation at 4x. We used the ELDA software package to interpret results (324).

2.3 Cloning, Plasmid Preparation & Infection

All plasmids were obtained via Addgene. The ClonTracer Library was a generous gift from the Stegmeier Lab (135), (plasmid #22240), the vector backbone is pRSI9-U6-(sh)UbiC-TagRFP-2A-Puro. pHIV-Zsgreen (Plasmid #18121), vector backbone is pSICO (6900bp without insert), was a gift from the Werb lab (328). pLM-vexGFP-Oct4 (plasmid #22240) was a generous gift from the Sadelain lab (329). Plasmid maps are shown in *Figure 2.3.0 A-C*. Packaging plasmids for production of lentiviral vectors were: pMDLg/pRRE (Plasmid #12251) containing *gag* and *pol* genes, pRSV-Rev (Plasmid #12253) containing *rev* gene (both were generous gifts of the Trono lab (330)) and pCMV-VSV-G (Plasmid #8454) containing envelope genes, a generous gift of the Weinberg lab (331). Packaging plasmid maps are shown in *Figure 2.3.0 D-F*.

2.3.1 Cloning

Plasmid cloning was carried out by the heat-shock method. 100ng stock plasmid was added to chemically competent *Escherichia coli* (*E. coli*) (One Shot TOP10, One Shot Stbl3, XL1-Blue (all Life Technologies) or Turbo Competent (New England Biolabs)) thawed on ice. Heat-shock was performed as per manufacturer's instructions at 42°C for 30 seconds. After recovery, cells were plated on appropriate antibiotic resistance plates (Luria broth (LB) agar with 0.1mg/ml ampicillin or 0.5 mg/ml puromycin) and single colonies picked for overnight culture at 37°C (Stbl3 at 34°C). Transformation (HIV-ZSG) of Stbl3 *E. coli* was found to result in significantly more transformants than TOP10, XL1-Blue or Turbo Competent (NED), and so was chosen for further studies (*Figure 2.3.1*). A control plate containing non-transformed Stbl3 *E. coli* was always included to ensure antibiotic selection effectiveness. Single colonies were selected from overnight cultures and expanded overnight in LB with 0.1mg/ml ampicillin (or 0.5 mg/ml puromycin) on orbital shakers (180 rpm, 37°C). Plasmids were purified using Zymo Plasmid Maxiprep kits according to manufacturer's instructions (alternative protocol). Purified plasmids

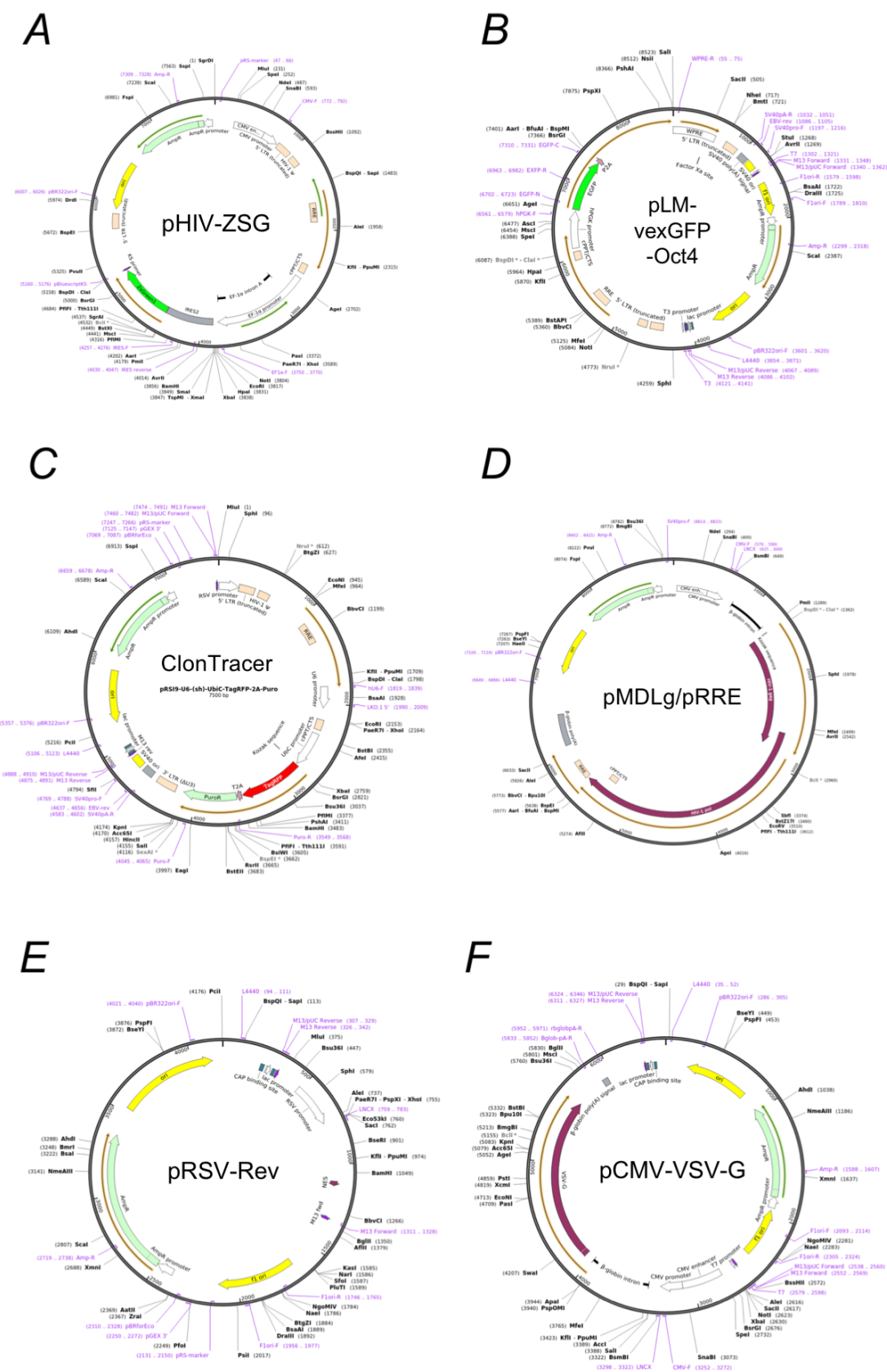
were DNA quantified by nanodrop, digested by restriction enzymes (standard protocols) and run on a 1% Agar gel in Tris base, acetic acid and EDTA (TAE) buffer with stock plasmid to visually confirm bands migrated at the same level (DNA stained with SybrSafe 1:10,000, Life Technologies).

ClonTracer Library expansion was as before but with modifications as follows: 10 vials of Stbl3 were heat shock transformed as normal and allowed to recover. Transformants were plated across 10x 30cm culture plates coated with LB agar & 0.5 mg/ml puromycin, giving a total *E. coli* culture area of 9,000cm². A 2µl sample was taken from each Stbl3 vial and plated on serial dilution plates (one at 1:1, 1:10, 1:100, 1:1000). Visual inspection of dilution plates enabled estimation of the total number of resistant colonies. If we estimated <2 million transformants across the 10 plates, then the experiment was discarded. Maxipreps were conducted directly from 30cm plate colonies without liquid culture. This process was repeated up to 5 times for lentiviral preparation. For single ClonTracer barcode preparations, individual bacterial colonies present on dilution plates with >1cm separation from neighbouring colonies were isolated and cultured for 3hrs in LB agar & 0.5 mg/ml puromycin on an orbital shaker (180 rpm, 37°C), bacteria were re-plated the process repeated to ensure single barcode colonies. Single barcode colonies were expanded in liquid LB broth & 0.5 mg/ml puromycin on an orbital shaker (24hrs, 180 rpm, 37°C), and plasmids were purified using Zymo Plasmid Maxiprep as before.

Figure 2.3.0 - Plasmid maps

*Figure 2.3.0 shows plasmid maps for plasmids obtained through Addgene. **A)** pHIV-ZsGreen1 (pHIV-ZSG), a generous gift from the Werb lab (328). **B)** pLM-vexGFP-Oct4, a generous gift from the Sadelain lab (329). **C)** ClonTracer Library (pRSI9-U6-(sh)-UbiC-TagRFP-2A-Puro), a generous gift from the Stegmeier Lab (135). **D)** pMDLg/pRRE containing gag and pol genes, a generous gift of the Trono lab (330). **E)** pRSV-Rev, containing rev gene, a generous gift of the Trono lab (330). **F)** pCMV-VSV-G containing envelope genes, a generous gift of the Weinberg lab (331). **Figure Overleaf.***

Figure 2.3.0 – Plasmid maps

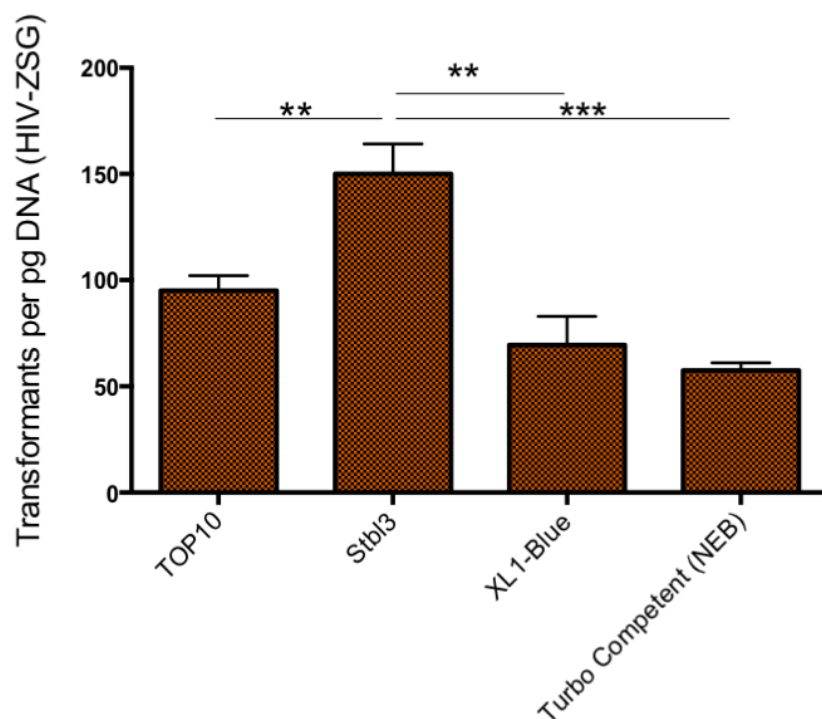


2.3.2 Lentiviral Packaging

Lentiviral plasmids were packaged into virions in HEK293-Ts of low (<20) passage, cultured in 10cm tissue culture dishes. Briefly, HEK293-Ts were culture in DMEM: F12 10% FBS with PenStrep as normal, trypsinised and plated at 2×10^6 /15 cm dish. Cells were cultured undisturbed overnight, aiming for 80% confluence. 1hr prior to transfection, media was removed and replaced by 5ml of DMEM:F12 + 5% FBS. Packaging plasmids (pMDLg/pRRE, pRSV-Rev and pCMV-VSV-G) were mixed with 3X insert plasmid (*e.g.* pHIV-Zsgreen) for a final concentration of 15µg /15cm dish (*i.e.* 2.5, 2.5, 2.5 and 7.5µg respectively). DNA was mixed with 250µl NaCl (150mM) and added to 40µl JetPEI (Polyplus) in 250µl NaCl (150mM). Mixture was incubated for 30mins at room temperature and added to each 15cm dish (dropwise; as per manufacturer's instructions). 5ml fresh media was added to each dish 2hrs post transfection. Media was collected and filtered through a 0.45µm filter unit after 24, 48 and 72hrs.

Viral supernatant (HEK media) was ultracentrifuged (SW28 rotor 2hrs 4°C @82700g) in 30ml Konical tubes (Beckman Coulter) with an underlay of 4ml sterile 20% sucrose solution ²⁵. Up to four sequential spins could be carried out in the same Konical tube to concentrate the virus (332). Viral pellets were resuspended in phospho-buffered saline (PBS) with 1% Bovine Serum Albumin (BSA) overnight at 4°C. Concentrated viral solution could be stored at -80°C until use. For ClonTracer viral preparation, this process was significantly scaled up, with up to 100x 15cm dishes in each viral prep (*Figure 2.3.2*).

²⁵ sucrose solution: 40g sucrose, 4ml 5M NaCl, 8ml 500mM HEPES pH 7.4, 0.4ml 0.5M EDTA, 170ml H₂O, filter sterilised.

Figure 2.3.1 – Competent *E. coli* comparison**Figure 2.3.1** – Competent *E. coli* comparison

Shows a comparison of commercially available competent *E. coli* preparations. Transforming Stbl3 cells resultant in significantly more drug resistant transformants per pg DNA. Statistical significance by one-way ANOVA with Bonferroni correction is denoted: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Data are presented as mean \pm standard deviation.

Figure 2.3.2 – *ClonTracer viral production*

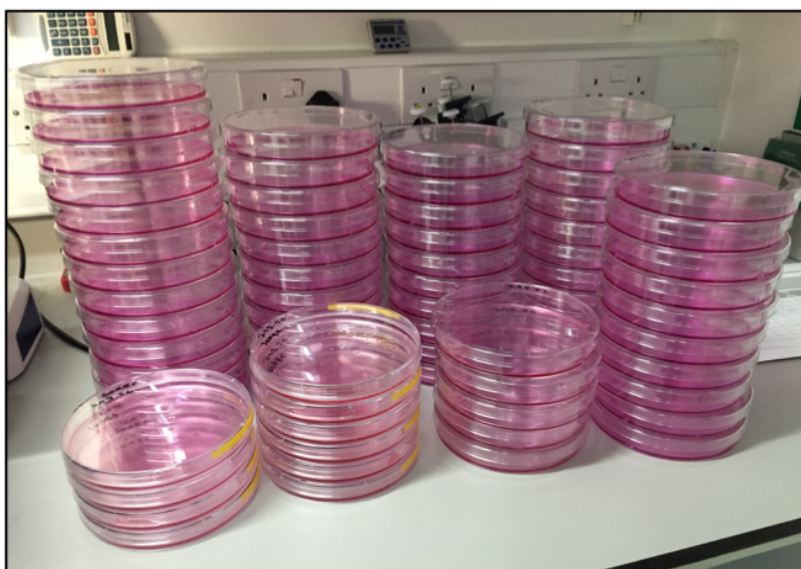


Figure 2.3.2 - *ClonTracer viral production.*

Figure 2.3.2 shows the viral production scale up for preparing ClonTracer lentiviral particles. HEK293-T cells are transformed with lentiviral constituent plasmids and supernatant (containing virions) is harvested over 72hrs and ultra-centrifuged.

2.3.3 Viral Titration and Multiplicity of Infection (MOI)

Lentiviral infection is thought to follow a Poisson distribution of equation:

$$P(n) = e^{-m} \cdot m^n / n!$$

Where, $P(n)$ is the fraction of cells infected by n virions and m is multiplicity of infection (MOI).

The distribution of fraction of cells with a given number of integrations versus the number of integrations ($P(n)$ vs n), given by this equation is shown in *Figure 2.3.3 A*. In our studies, a of 0.1 was targeted to ensure that most cells with any integration events ($\sim 10\%$) had a single integration event (41). Various methods exist to estimate virus titer, the two most common are direct p24 measurements via enzyme-linked immunosorbent assay (ELISA) and qPCR for viral RNA. Each can overestimate the number of infectious virions in a preparation as the measurement will include defective particles and/or unintegrated viral particles. Functional titres calculate Transforming Units (TU), *i.e.* functional virions, as shown for HIV-ZSG in *Figure 2.3.3 B*.

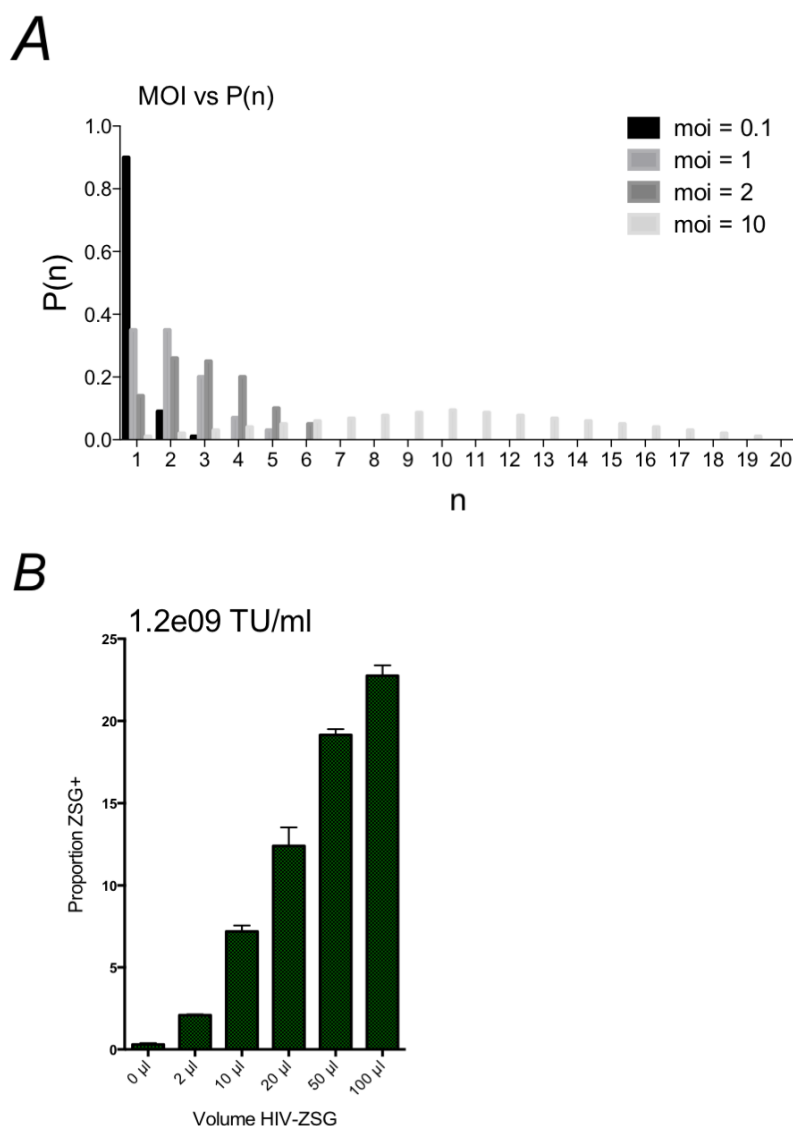
Figure 2.3.3 – Multiplicity of infection (MOI) calculations**Figure 2.3.3 - Multiplicity of infection (MOI) Calculations.**

Figure 2.3.3 shows calculation of the MOI. **A)** Shows the distribution of cells with a given number of integrations ($P(n)$ vs n), for various different MOIs. given by the equation: $P(n) = e^{-m} \cdot m^n / n!$ **B)** Shows calculation of virions present in a HIV_ZSG viral preparation, using HEK293-T cells we find 1.2×10^9 transforming units (TUs) per ml. **Figure Overleaf.**

2.3.4 Lentiviral Infection

PDTCs were generated from PDX tumours as described and cultured in MEGM in ultra-low attachment plates with B27 –vitamin A added to the growth media. MCF7 and other established cell lines were cultured as described in normal growth media until ~70% confluence was achieved. Immediately on plating, concentrated virus (equivalent to 5 x 10cm dish per well of 6-well plate) was added to PDTCs with up to 10µg/ml Hexadimethrine bromide (Polybrene). Maximal Polybrene concentration for each model was determined as in *Figure 2.3.4* (showing STG282x4 PDTCs as an example). For cell lines, virus was added at appropriate MOI with up to 10µg/ml Hexadimethrine bromide, media was changed after 4hrs. For PDTCs, virus was not removed during the infection process, as plates would require centrifugation to change media. PDTCs were cultured in the presence of lentivirus for 24-48hrs before flow cytometry, cell sorting or puromycin selection.

2.3.5 Puromycin Selection

Puromycin toxicity was determined in MCF7s using the same protocol as for 4-OHT dose response analysis. We identified toxicity (defined as <1% viability after 48hr culture) of MCF7 cells through puromycin titration to be between 1,000ng/ml and 2,000ng/ml, in line with values found in the literature (43), and so used 2,000ng/ml for 48hrs in selection of infected MCF7s. Several PDTC models were found to be resistant to puromycin at concentrations up to 10,000ng/ml. To minimise any potential off target effects of puromycin in these cells, we opted for flow cytometry based viral titration.

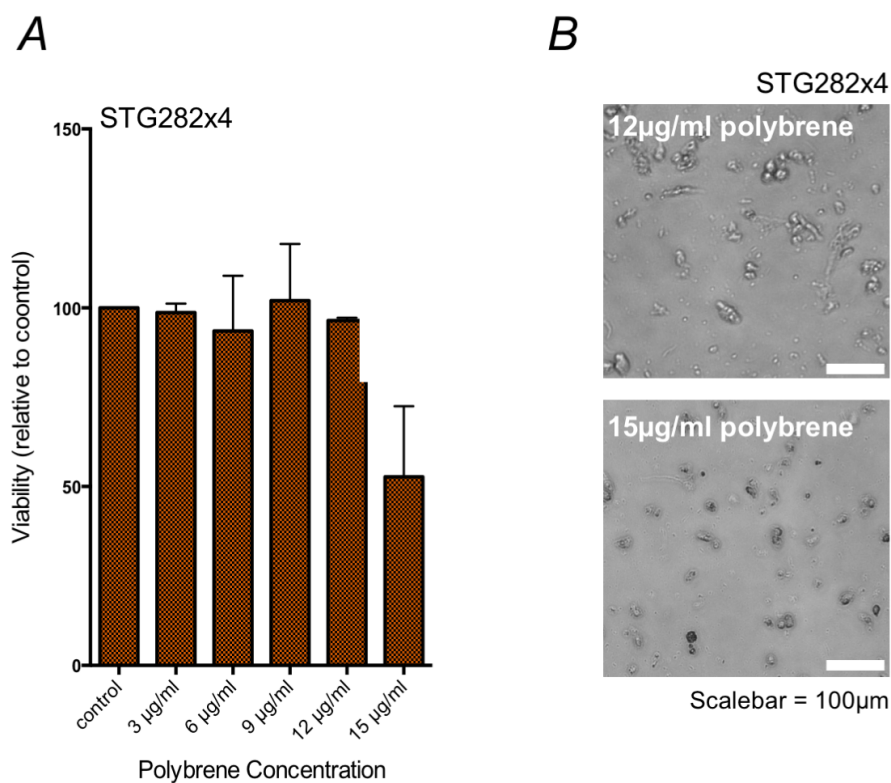
Figure 2.3.4 – Polybrene optimisation**Figure 2.3.4** - Polybrene optimisation.

Figure 2.3.4 shows a calculation of effective polybrene concentration for aiding in infection of PDTs. **A)** Shows viability of STG282x4 at increasing concentrations of polybrene using a CellTiter-Glo Assay. **B)** Shows phase contrast microscopy of the same STG282x4 cells after 3 days in culture with polybrene, immediately before CellTiter-Glo Assay.

2.3.6 Flow Cytometry and Cell Sorting

24-48hrs post infection, PDTCs were mechanically dissociated into single cells. Media was removed and PDTCS were blocked in HBSS with 5% normal Rat Serum (Sigma) at 4°C for 30 minutes. Cell lines were harvested as normal. Cells were washed in MEGM and filtered through a 4µm cell strainer. For ALDH activity assays, the ALDEFLUOR Kit (Stemcell Technologies) was used as per manufacturer's instructions. For cell sorting experiments, cells were transferred to FACs tubes and stained with pan mouse MHC-I/MHC-II (H-2k^b/H2-D^b) conjugated fluorescein isothiocyanate (FITC) (Biolegend, Mouse C3H IgG2_{a,k})²⁶. After 30 minutes, 0.5µl DAPI was added for live/dead discrimination and PDTCs were analysed by flow cytometry using standard excitation and emission spectra for green fluorescent protein (GFP; in place of specific optimisation for ZsGreen1) or red fluorescent protein (RFP). To correct for spill-over, spectral overlap values were measured for all fluorophores and in all detectors, via single-color controls. The spill-over values were then used to calculate compensation values by FlowJo 10.5.0 (for MacOS). Cells were gated and sorted if required using standard protocols for the FACSaria. After flow sorting, cells were collected in MEGM with 20% PenStrep and washed three times in MEGM supplemented with 20% PenStrep (Life Technologies) to minimise the possibility of contamination from the flow cytometer. For mouse studies, around 1 x10⁵ cells were resuspended in 100µl and 100µl Matrigel was added immediately prior to injection into the flank of 21-day-old NSG mice²⁷. Cells were analysed and sorted using the BD FACSaria or the BD Influx FACS machine with data files analysed using FlowJo 10.5.0 (for MacOS). The CI Flow Cytometry Core aided in the completion of these experiments.

²⁶ H-2k^b/H2-D^b has been validated in our lab as 93% sensitivity and 92% specificity for detecting host stromal cells in PDX models by using mouse whole blood as a positive control (198).

²⁷ Matrigel used in these experiments were of the same batch /lot number. However, it should be noted that routine passage of the PDX biobank did not necessarily use the same Matrigel batch, which could lead to variability between out experimental models and the PDX biobank.

2.3.7 siRNA transfection

Short interfering RNAs (siRNAs) for *POU5F1* (OCT4: Chr.6: 31164337 - 31170693 on Build GRCh38, catalogue #AM16708) and *SOX2* (SOX2: Chr.3: 181711924 - 181714436 on Build GRCh38, catalogue #106971) were reverse transfected into MCF7 cells with the Stealth siRNA Lipofectamine RNAiMAX kit (Invitrogen). MCF7 cells of <10 passage were maintained in D-MEM/F12 supplemented with 10% heat-inactivated foetal bovine serum, 2mM glutamine, and penicillin/streptomycin and passaged into 24-well plates. 6pM siRNA was diluted in 100µl Opti-MEM I Medium without serum in each transfection well. 1µl Lipofectamine RNAiMAX was added and mixture incubated for 25 minutes at ambient temperature. 50,000 cells in 500µl complete growth media minus antibiotics was added to each well containing siRNA + lipofectamine (final siRNA concentration 10nM) and wells plated rocked gently for 5 minutes at ambient temperature. Cells were then incubated for 48hrs at 37°C in a 5% CO₂ incubator until ready to assay for gene knockdown. Control cells were incubated with scrambled control siRNAs supplied in the siRNA kit (Invitrogen). Efficient knockdowns were ensured by western blot.

2.3.8 Gel Electrophoresis, extraction and purification

Analysis by gel electrophoresis was performed by adding 5µl of digestion reaction together with 1% 10x loading dye into the wells of a homemade 1% agarose Tris base, acetic acid and EDTA (TAE) gel containing a 1:10,000 dilution of SYBR safe DNA gel stain (Invitrogen) and run at 75nV. A 100bp or 1kb Quick-load DNA ladder (NEB) was used as appropriate. As soon as DNA fragments were clearly separated, gel sections containing the desired fragments were extracted and purified using the QIAquick Gel Extraction kit (Qiagen), following the manufacturers recommended protocol. DNA was eluted in 50µl of nuclease-free water (Sigma).

2.4 Molecular Profiling

2.4.1 Real-time quantitative polymerase chain reaction (RT-qPCR)

RNA extraction was carried out using a Tissue Homogenising Kit (CK28-R, Precellys) and miRNeasy Mini Kit (Qiagen) according to manufacturer's instructions (tissue extraction protocol). Genomic DNA was removed from RNA samples prior to RT-PCR using the Ambion DNase Treatment and Removal kit (Applied Biosystems), according to manufacturer's instructions (rigorous treatment protocol). RNA was quantified using a NanoDrop instrument (Thermo Scientific). Total RNA (1µg) was reverse transcribed to cDNA using a Transcriptor First-Strand cDNA synthesis kit (Roche) in duplicate in a 20µl reaction mixture as per manufacturer's instructions (oligo DTs and random primers or TaqMan miRNA specific primers as appropriate). qPCR was carried out using TaqMan predesigned probes (Life Technologies; as per manufacturer's instructions) and TaqMan Fast Universal PCR master mix (2x) (Applied Biosystems) in an Applied Biosystems 7900HT thermocycler (fast 364-well protocol)²⁸. A melting curve analysis was performed from 70°C to 95°C in 0.3°C intervals to demonstrate the specificity of each amplicon and to identify the formation of primer dimers. The geometric mean of C_T values for glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and β2-Microglobulin (B2M) was used as a reference gene to normalise all data (internal control). RQ (relative gene expression) values were calculated using a comparative C_T method (also known as delta delta C_T method or 2^{-ΔΔC_T}) (333). TaqMan probes used were as shown in *Table 2.4.1*.

²⁸ Conditions were as follows: 95°C for 15 mins, followed by 40 cycles of three steps consisting of denaturation at 94°C for 15 secs, primer annealing at the optimal temperature for 30 secs and primer extension at 72°C for 30 secs.

Gene Name	ProbeID	Dye
B2M	Hs00187842_m1	VIC-MGB
CDH1	Hs01023895_m1	FAM-MGB
CDH2	Hs00983056_m1	FAM-MGB
CLDN3	Hs00265816_s1	FAM-MGB
CLND5	Hs00533949_s1	FAM-MGB
ESR1	Hs01046816_m1	FAM-MGB
GAPDH	Hs02786624_g1	VIC-MGB
GREB1	Hs00536409_m1	FAM-MGB
PGR	Hs01556702_m1	FAM-MGB
POU5F1	Hs04260367_gH	FAM-MGB
PROCR	Hs00197387_m1	FAM-MGB
SOX2	Hs01053049_s1	FAM-MGB
TGFB1	Hs00998133_m1	FAM-MGB
TWIST1	Hs01675818_s1	FAM-MGB
VIM	Hs00958111_m1	FAM-MGB
WISP2	Hs01031984_m1	FAM-MGB

Table 2.4.1 – qPCR TaqMan probes from Thermo-Fischer

2.4.2 Western Blotting

Cells from fresh PDTC or cell line preparations were trypsinised to make a single cell suspension and washed in PBS and pelleted in 1.5ml Eppendorf tubes. Whole cell lysates or cryopreserved cell lines /PDTCs previously prepared and stored in Serum and 10% DMSO were defrosted on ice, vortexed for 20 seconds and centrifuged for 20 minutes at 14,000 rpm (4°C). Fresh tissue was washed in PBS, homogenised using the Tissue Homogenising Kit (CK28-R, Precellys) and trypsinised to make a single cell mixture. Each preparation was treated with between 50-150µl complete lysis buffer ²⁹ (depending on size of pellet), mixed thoroughly by pipetting and incubated for 30 minutes at 37°C. Pierce BSA protein quantification assay was carried out as per manufacturer's

²⁹ 5ml 1M Tris-HCL pH 7.5 (=50mM), 1ml 0.5M EDTA (=5mM), 10ml 1.5M NaCl (=150mM), 1ml TritonX-100 (=1%), 5ml 1M NaF (=50mM), 2.5ml 1M β-glycerophosphate (=25mM). Mixture was made up to 100ml with dH₂O (75.5ml) and 1x Complete EDTA-free protease inhibitor cocktail tablet (Roche) and 5µl Sodium orthovanadate 1M (Na₃VO₄) added per 10ml of protein lysis buffer. Aliquots could be stored for up to 1 month at -20°C.

instructions (Thermo Scientific) with an EnVision plate reader (PerkinElmer) and standard curve prepared in GraphPad Prism 5 (for MacOS). 5µl loading buffer (Laemmli 2x concentrate (Sigma Aldrich) was added to 20µg protein from each sample and mixture boiled. Laemmli buffer contains sodium dodecyl sulphate (SDS) to negatively charge proteins and 2-mercaptoethanol to reduce disulphide bridges. Samples were loaded and separated by gel electrophoresis ³⁰. Protein were transferred to a nitrocellulose membrane using an iBlot transfer device and iBlot transfer stacks (Invitrogen). Membrane was blocked overnight (LI-COR Odyssey blocking buffer) and stained with primary (Anti-SOX2 antibody (Abcam, ab97959), Anti-Oct4 antibody - ChIP Grade (Abcam, ab19857)) and secondary (LI-COR) antibodies as per manufacturers recommendations.

2.4.3 Immunofluorescence and Confocal Microscopy

Monolayers were cultured as normal in Falcon Culture Slides (Fisher) and mechanically scratched when 80% confluent using a p10 pipette tip. 24hrs later, cells were fixed with 4% paraformaldehyde for 30 minutes at room temperature. Cells were washed with PBS and blocked in HBSS with 10% Normal Rat Serum for a further 3hrs. Cells were washed with PBS and conjugated northern lights fluorescent antibodies for E-cadherin, Vimentin and SNAI1 were added at 1:250 (3 colour EMT kit, R&D Systems). Cells were incubated with antibodies for 24hrs at 4°C with gentle rocking and protection from light. After 24hrs, stained cells were washed in PBS and images were acquired using a Leica Tandem Confocal system with appropriate excitation strengths and filters optimised for northern lights conjugated antibodies. Microscope settings were never changed between images.

³⁰ Conditions were: Nu-Page tris glycine polyacramide gel 10-20%; 2-(N-morpholino) ethanesulfonic acid (MES) running buffer; Invitrogen), LI-COR molecular weight ladder was included.

2.4.4 Immunohistochemistry and Histology

Immunohistochemistry (IHC) was carried out by the histopathology core facilities at the CRUK Cambridge Institute using standard procedures and protocols. Dr Hamid Raza Ali, a breast cancer pathologist, aided in the interpretation of data and optimised staining procedures for antibodies used. HER2 staining was performed at the Department of Histopathology at the Addenbrooke's Hospital using the diagnostic standard antibody, comprised of a Dako autostainer and the Herceptest kit (Dako).

Paraffin-embedded tissue blocks were sectioned, and sections dewaxed by 2x 10 minutes immersions in xylene (Fischer Scientific). Rehydration was achieved by 2x 5 minutes immersions in 100% ethanol (EtOH) followed by a 5 minute emersion in 70% EtOH (Fischer Scientific). Dewaxing and hydration were conducted on the automated Leica ST5020 multistainer system (Leica Microsystems). Heat-induced antigen retrieval (with sodium citrate) was performed at 100°C and protease-induced antigen retrieval (with Bond enzyme concentrate (Leica Microsystems)) at 37°C. IHC was performed using the Leica BOND-MAX autostainer (Leica Microsystems). Bound primary antibodies were detected using a horseradish peroxidase (HRP) polymer-conjugated secondary antibody or a biotinylated secondary antibody (both from Bond Polymer Refine Detection Kit; Leica Microsystems). In the first case, primary antibodies raised in mouse or rabbit are detected by a secondary linker anti-mouse antibody raised in rabbit (or *vice versa*), followed by detection of the linker antibody by the HRP polymer-conjugated secondary antibody. Binding of the secondary is visualised using 3,3-diaminobenzidine tetrahydrochloride (DAB), which is a substrate for the HRP enzyme. Staining was intensified using DAB enhancer (Leica Microsystems). In the case of Polymer Refine Detection, biotinylated secondary antibodies are used to detect bound primary antibody. The biotinylated secondary is then detected by HRP-streptavidin conjugated antibody and DAB staining performed as described. Details of antibodies used are detailed in *Table 2.4.4.1* (primary) and *Table 2.4.4.2* (secondary).

Protein	Clonality	Source	Catalogue #	Dilution	Retrieval
ALDH1a	Rabbit Polyclonal	Atlas	HPA002123	1:275	Sodium Citrate, 20'
CD24	Mouse Polyclonal	Thermo Fischer	MA5-11833	1:200	Sodium Citrate, 30'
CD44	Rabbit Polyclonal	Atlas	HPA005785	1:800	Sodium Citrate, 30'
E-Cadherin	Mouse Polyclonal	Dako	M3612	1:25	Sodium Citrate, 30'
N-Cadherin	Mouse Polyclonal	AbCam	ab18203	1:250	Sodium Citrate, 20'
EpCam	Mouse Polyclonal	Novocastra	NCL-ESA	1:100	Proteolytic Enzyme (1:167), 10'
ER	Mouse Polyclonal	Novocastra	NCL-ER-6F11/2	1:70	Sodium Citrate, 30'
p53	Mouse Polyclonal	Dako	M7001	1:1000	Sodium Citrate, 30'
Vimentin	Mouse Polyclonal	Novocastra	NCL-L-VIM-572	1:400	Tris EDTA, 20'
PR	Mouse Polyclonal	Dako	M3569	1:50	Sodium Citrate, 30'
Ki67	Mouse Polyclonal	Dako	M7240	1:200	Tris EDTA, 30'
GFP	Chicken Polyclonal	AbCam	ab13970	1:100	Proteolytic Enzyme (1:333), 20'

Table 2.4.4.1 – details of primary antibodies used in IHC

Antibody	Clonality	Source	Catalogue #	Dilution
Biotinylated anti-rabbit	Donkey polyclonal	Jackson	711-065-152	1:250
Biotinylated anti-rat	Donkey polyclonal	Jackson	712-065-153	1:250
Biotinylated anti-chicken	Donkey polyclonal	Jackson	703-066-155	1:250

Table 2.4.4.2 – details of secondary antibodies used in IHC

2.5 Barcode Analysis

2.5.1 Next Generation Barcode Sequencing & Analysis

PCR-amplified products were quantified using the Standard Sensitivity NGS Fragment Analysis Kit (Advanced Analytical Technologies) on the Advanced Analytical Fragment Analyzer Automated CE System. Barcode PCR reaction conditions were as follows: per 50µl reaction, PCR-grade water up to 50µl, 10x Titanium Taq PCR buffer 5µl, 50x dNTP mix (10mM each) 1µl, F Seq primer (10µM) 1µl, R Seq primer (10µM) 1µl, 50x Titanium Taq DNA polymerase (CLONTECH) 1µl, DMSO 2µl, gDNA template (2µg per reaction)³¹. Fragment sizes and concentrations of PCR amplicons were analysed using Advanced Analytical PROSize 2.0 software. PCR amplicons were then prepared at 10nM and loaded at 3.5pM for sequencing on the Illumina HiSeq2500 sequencer in Rapid Mode using the 50 Cycle TruSeq Rapid SBS Kit, TrueSeq Rapid SR Cluster Kit, and HiSeq Rapid SR Flow Cell (Illumina). Barcode Primers are shown in *Table 2.2.5.1*.

³¹ Cycling conditions: 95°C (5 minutes; 1 cycle), 35 cycles of 95°C (30 seconds) 66°C (30 seconds) 72°C (1 minute), 1 cycle of 72°C (7 minutes). Cycling conditions, including cycle numbers, followed the protocol developed by *Bhang et al.*, (334).

WS PCR Forward Primer	Sequence	Length
	AATGATACGGCGACCACCGAGATCTACACACTGACTGCAGTCTGAGTCTGACAG	54

Barcode PCR Reverse Primer	Barcode PCR Reverse Primer Sequence	Length
WS_Rev_Index_001	CAAGCAGAAGACGGCATACGAGATACGATCGTGAAGTCTGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_002	CAAGCAGAAGACGGCATACGAGATCTAGATCGTGGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_003	CAAGCAGAAGACGGCATACGAGATGACTCGATCAGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_004	CAAGCAGAAGACGGCATACGAGATTGACTAGCTCGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_005	CAAGCAGAAGACGGCATACGAGATATGCTCAGCAGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_006	CAAGCAGAAGACGGCATACGAGATCGATCTGCATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_007	CAAGCAGAAGACGGCATACGAGATGATAGCTGACTGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_008	CAAGCAGAAGACGGCATACGAGATTCAAGTACGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_009	CAAGCAGAAGACGGCATACGAGATAGTACGCATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_010	CAAGCAGAAGACGGCATACGAGTACGATCGATAGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_011	CAAGCAGAAGACGGCATACGAGATTGATCAGCAGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_012	CAAGCAGAAGACGGCATACGAGATCGCAGTACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_013	CAAGCAGAAGACGGCATACGAGATAGCGTCTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_014	CAAGCAGAAGACGGCATACGAGATCAGCATGTCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_015	CAAGCAGAAGACGGCATACGAGATTACTCATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_016	CAAGCAGAAGACGGCATACGAGATTCTGCAGCTAGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_017	CAAGCAGAAGACGGCATACGAGATACTGTACTCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_018	CAAGCAGAAGACGGCATACGAGATCGACAGCTATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_019	CAAGCAGAAGACGGCATACGAGATGTATCGGTAGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_020	CAAGCAGAAGACGGCATACGAGATTAGTCGCATGTTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_021	CAAGCAGAAGACGGCATACGAGATATCGATGACGGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_022	CAAGCAGAAGACGGCATACGAGATCGATAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_023	CAAGCAGAAGACGGCATACGAGATGAGCTGTATCGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_024	CAAGCAGAAGACGGCATACGAGATTCTGATCGCAGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_025	CAAGCAGAAGACGGCATACGAGATAGCATCGTCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_026	CAAGCAGAAGACGGCATACGAGATCTACGTCTAGGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_027	CAAGCAGAAGACGGCATACGAGATGCTAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_028	CAAGCAGAAGACGGCATACGAGATTGAGTGCATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_029	CAAGCAGAAGACGGCATACGAGATACGCTGACATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_030	CAAGCAGAAGACGGCATACGAGATCATACAGTGCCTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_031	CAAGCAGAAGACGGCATACGAGATGAGCACTAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_032	CAAGCAGAAGACGGCATACGAGATTGCATGTAGCGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_033	CAAGCAGAAGACGGCATACGAGATAGTGATCGAGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_034	CAAGCAGAAGACGGCATACGAGATCTGACATGCAGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_035	CAAGCAGAAGACGGCATACGAGATTGACAGATCGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_036	CAAGCAGAAGACGGCATACGAGATTCACTATGCGGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_037	CAAGCAGAAGACGGCATACGAGATCTCGATACGGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_038	CAAGCAGAAGACGGCATACGAGATCGCATGATGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_039	CAAGCAGAAGACGGCATACGAGATGACAGTCACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_040	CAAGCAGAAGACGGCATACGAGATTGCACTAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_041	CAAGCAGAAGACGGCATACGAGATATCAGCGATGTTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_042	CAAGCAGAAGACGGCATACGAGATCTGTATGAGCGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_043	CAAGCAGAAGACGGCATACGAGATTGACTGTCACTGTTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_044	CAAGCAGAAGACGGCATACGAGATTACGCTGCATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_045	CAAGCAGAAGACGGCATACGAGATAGCTGATGCACTGTTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_046	CAAGCAGAAGACGGCATACGAGATCTATGCATGTTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_047	CAAGCAGAAGACGGCATACGAGATGCTCATGTCACTGTTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_048	CAAGCAGAAGACGGCATACGAGATTAGCGATCTGTTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_049	CAAGCAGAAGACGGCATACGAGATACGTACTGCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94
WS_Rev_Index_050	CAAGCAGAAGACGGCATACGAGATCATAGCATCGTGAAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCCTAGCT	94

Table 2.5.1 – Primers used for barcode library preparation.

Table shows primers used for cell lines and PDX models in ClonTracer studies. Forward primer spans 54bp upstream of the barcode insert, and forms Illumina P5 adapter of amplicon. Reverse primer spans 94bp downstream of the barcode insert and forms an inline barcode (for demultiplexing) and the P7 adapter of the resultant amplicon.

2.5.2 Barcode Composition Analysis

FASTQ files were prepared from the barcode-sequencing runs to count the number of reads and the fraction of barcodes in each sample. Demultiplexing was performed using the FastX toolkit (adapter trimming function) allowing for up to one mismatch (http://hannonlab.cshl.edu/fastx_toolkit/). All further analysis was undertaken in python using the ClonTracer Analysis package (Algorithms for Deconvoluting NGS data V1.2) (<https://bit.ly/2Lv5UVs>) (135). Reads were then filtered to keep those that (a) showed the WS × 15 pattern; (b) matched the expected sequence after the WS × 15 barcode for sequence libraries with lengths of ≥37 bp; and (c) had an estimated Phred quality score of at least 10 for all base pairs in the read, with an average Phred quality score greater than 30. All barcodes observed at least twice that passed these criteria were kept. Sets of barcodes were then merged to account for sequencing errors if either (a) the test barcode was a hamming distance of 1 from the more abundant barcode and observed at 1/8th of the count or (b) the test barcode was a hamming distance of 2 from the more abundant barcode and observed at 1/40th the count; this is similar to the approach used by *Lu et al.*, (202) and *Bhang et al.*, (334). For those sequence libraries with lengths of exactly 30 bp, the hamming distance also allowed insertions or deletions of 2 bp; because of the barcode design and our filtering steps, insertions or deletions of odd-numbered length or in the longer reads should already have been filtered out. After these steps, each barcode set was annotated with the sequence of the most abundant barcode, counts before merging, counts after merging, and fraction with respect to the total count of all barcodes that passed the read filters.

2.5.3 Calculation of Barcode Overlap Between Samples

To establish the significance of sharing between samples treated with different drugs, we computed 'sharing ratios'. The sharing ratio was defined as the ratio of significant barcodes of one replicate seen in another to the total number of significant barcodes in that replicate. A barcode was called significant if it was seen in 0.021% of the total (which

was the highest fraction observed in EtOH-treated groups). A barcode was considered shared between two samples if it was significant in both samples and, to rule out possible errors due to low-level contamination during sequencing, the fraction of the barcode in the less abundant sample was at least 1% compared to the fraction in the more abundant sample.

2.5.4 Spike-in Experiments

HEK293-T cells were infected at a low MOI (<0.1) with one of 10 clonally derived ClonTracer barcodes, prepared as lentiviral constructs as normal. Known numbers of these individually-tagged cells, spanning a range of 250 to 10^5 cells, were spiked into each PDX sequencing run. Genomic DNA was extracted from single barcoded HEK293-T cells samples, mixed with degenerate barcoded PDX cell samples. In each PDX replicate, up to four separate HEK293-T dilutions were included. Additionally, one well per sequencing run contained 10 dilutions of between 250 to 10^5 cells. From combined data across each sequencing run, we therefore had 116 known dilutions in total. The sensitivity and specificity of clone detection for each input cell number (*i.e.* clone size) could thus be calculated. These internal controls allowed us to compute the estimated clone size for each calculated barcode abundance. We found 1,000 cells or above had a sensitivity and specificity of 100%, whilst 500 cells had a sensitivity of ~60% and specificity of ~95% (one false positive clone (out of 20) fitting barcode detection and filtering criteria but not present in spike-ins). Our limit of accurate detection was thus set at 500 cells, the relative barcode abundance (RBA) equivalent of which, varied slightly between test samples. Thus, these spike-ins acted as an internal control, allowing normalisation between barcode analysis of different PDX models.

2.6 Next Generation Sequencing

NGS libraries were prepared by the CI genomics core, with bioinformatics (VAF filtering and PyClone) run by Dr Oscar Rueda, using protocols defined in (198).

2.6.1 Whole Exome Sequencing

WES libraries were prepared using Nextera Rapid Capture Exome (Illumina Inc., USA) following manufacturer's instructions (Enrichment Guide version #15037436 Rev. J, Illumina Inc., USA). Pre-capture libraries were quantified using qPCR and their average length was assed using DNA1000 chip on Bioanalyzer 2100 (Agilent Technologies Inc., USA). 500ng of each library was pooled for three-plex exome capture, and after 11 cycles of PCR amplification were again normalized to 15nM and pooled for high-coverage paired-end exome sequencing. The sequencing was performed using 125bp paired-end reads. Bam files were merged, sorted and indexed using samtools. Duplicates were marked using Picard tools and insertions and deletions (indels) were realigned using GATK. HaplotypeCaller was employed for variant calling, and after that several filters were applied using the Bioconductor package VariantAnnotation: for single nucleotide variants (SNVs), a minimum genotyping quality of 20, at least 5 reads at the variant position, a strand bias Phred-scale p value smaller than 40 and no presence of homopolymers in the surrounding region. For indels, we increased the width of the region to detect nearby homopolymers. Genotypes and variant allele frequencies (VAFs) were computed from these calls.

2.6.2 Shallow Whole Genome (sWGS)

As described above in exome sequencing analysis, libraries were prepared using Nextera Rapid Capture Exome (Illumina Inc., USA). 50 bp single-read whole-genome shallow sequencing was performed in parallel with the exome sequencing to provide a clean and

accurate estimate of copy number. Alignment was performed using BWA and data were analysed using the Bioconductor package QDNaseq. Copynumber (HOMD, Homozygous deletions, HETD, Heterozygous deletions, NEUT, neutral copy number, GAIN, single copy gains and AMP, high-level amplifications), were called based on thresholds on the segmented mean log₂-ratio (-1, -0.4, 0.25, 0.75).

2.6.3 PyClone Clonal Reconstruction

PyClone takes as input the allele frequencies of somatic mutations in each sample and clusters mutations that shift together across the samples, predicting the cellular frequency for each cluster in each sample accounting for copy number changes and normal cell contamination. PyClone was run in Python using recommended settings for 40,000 iterations with a burn in period of 20,000 iterations using a beta binomial parameter of 500 (335). To measure the clones whose prevalence changed significantly between time points, we compute the 90% credible intervals and those clusters that had a sample not overlapping with the rest were called significant.

2.7 Statistical Analysis

Data were tested for Gaussian distribution by the Shapiro-Wilk Test with a cut-off of 0.05. For normally distributed data, comparisons between two groups were made by the 2-sample t-test and more than two groups by one-way analysis of variance (ANOVA) with Bonferroni correction. All t-tests were two tailed. For non-Gaussian data, non-parametric comparisons were made by Mann-Whitney or Kruskal-Wallis for multiple comparisons. All statistical and graphical analysis of data was carried out in GraphPad Prism 6.0 (for MacOS). Statistical significance is denoted where appropriate (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Statistical significance was only calculated between groups of at least 3 biological replicates and if possible where technical replicates were available to act as controls (10% variability between technical replicates was considered acceptable and means were taken forward as biological replicates). All data is presented as mean \pm standard deviation, unless otherwise noted.

Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (7) and The Cancer Genome Atlas (TCGA) Breast Cancer (336) datasets were downloaded from cBioportal (337,338) and the TCGA data portal (tcga-data.nci.nih.gov). Normalised RNASeqV2 expression values and corresponding clinical metadata were manipulated in R Studio (Mac) 0.99.484 (R Studio: <https://www.rstudio.com/>). Combined data were analysed in Microsoft Excel (Mac 14.4.3) and R Studio with results plotted in GraphPad Prism 6 (Mac). Comparisons between two groups were made using two-tailed t-tests or Mann-Whitney U-tests and more than two groups by one-way ANOVA or Kruskal-Wallis, multiple comparisons were corrected for by the Bonferroni method. Kaplan-Meier disease free survival graphs were constructed in GraphPad Prism 6 and presented as whole population versus experimental population. Comparisons between curves were made by Log-rank (Mantel-Cox) tests with the Mantel-Cox p value reported.

Chapter 3

Tracking Tamoxifen Resistance with High Complexity Barcodes

Aims of the chapter

1. Establish a reproducible method for induction of tamoxifen resistance in breast cancer MCF7 cell line.
2. Characterise and expand a high complexity barcode library and label a population of parental MCF7s.
3. Induce resistance in barcoded MCF7s and follow each cell's clonal fate by NGS of barcodes.
4. Determine whether the development of resistance in this system is *de novo* or results from clonal expansion of a specific resistant population.

3.1 Introduction

Despite remarkable advances in our understanding of the progression of human malignancies and the molecular events that underpin tumour survival, new therapies often fail to show significant efficacy in clinical trials. Even with patient stratification, clinical responses can be fleeting, often adding only 6-12 months before disease progression (339). As discussed, tumours have long been known to be heterogeneous in nature and composed of multiple cellular phenotypes (340). This intratumour heterogeneity is governed by both cell-autonomous (*e.g.* genomic and epigenomic heterogeneity) and non-cell-autonomous (*e.g.* stromal heterogeneity) factors (124) and has clinical implications in patient specific responses to therapy and the rapid emergence of drug resistance (134). In this chapter, we set out to elucidate the role of intratumour heterogeneity in the development of resistance to a clinically relevant agent in breast cancer, tamoxifen.

ER α is the defining feature of the majority of breast cancers and is thought to be a key mediator of proliferation, differentiation and apoptosis in ER α + tumours and the developing mammary gland (341). Tamoxifen, an ER α antagonist (via its metabolite 4-hydroxytamoxifen (4-OHT)) is widely used as an anti-oestrogen therapy for both early and late stage ER α -positive (ER α +) breast cancer in pre- and post-menopausal women (255). Despite significant benefit to the majority of ER α + patients, both innate and acquired resistance to tamoxifen are common clinical outcomes (255). Tamoxifen resistance in ER α + tumours is generally associated with differential ER α -binding programmes mediated by FOXA1, increased *AIB-1* expression and loss of *ERBB2*, repression by the ER α -PAX2 complex (248,272,273). Along with large scale transcription factor binding studies in patient samples, the breast cancer cell line MCF7 has been used extensively to study the acquisition of tamoxifen resistance (272,342,343).

The emergence of tamoxifen resistance in the ER α and progesterone receptor (PR) positive (HER2 negative), luminal epithelial cell line MCF7 is a well-documented process (343,344). As the development of resistance is a predictable outcome of treating the parental cell line with low doses of tamoxifen, it is possible that a drug resistant clone exists in the population prior to treatment (135). MCF7s are amenable to extremely high throughput culture with few technical requirements. Prolonged culture of MCF7s in the presence of 4-OHT, or in the absence of exogenous oestrogens, leads to the clonal outgrowth of resistant cellular populations. Sub-lines are heterogeneous in nature but are typically characterised by rapamycin resistance and increased phosphorylation of Akt/mTOR pathway members (345). Sub-lines developed in the absence of oestrogens are further characterised by the acquisition of *PAX2* expression and loss of active phospho-HER2 (345). Interestingly, there is evidence for the role of the transcription factor SOX2 in MCF7s selected for tamoxifen resistance ('TamR' cell lines), through supporting a stem/progenitor cell population (342,346). SOX2 is best known for its role in maintaining pluripotency in embryonic stem cells (ES cells) and as one of four transcription factors required to induce pluripotency in terminally differentiated cells (induced pluripotent stem cell; iPS cells) (280). Additionally, its role in regulating self-renewal and tumourigenicity of melanoma-initiating cells has recently been described (278).

Typically, in cancer cell line models, resistance to targeted therapies is thought to arise from pre-existing populations, rather than from *de novo* evolution (135). However, these clonal populations can be defined by phenotypic as well as genomic traits. In an elegant study by Kreso *et al.*, it was found that minor 'Type IV' subclones in colorectal PDXs were able to repopulate the tumour bulk after treatment with chemotherapy (196). These quiescent cell populations were genetically similar to their highly proliferative counterparts, and were later linked to the BMI1 positive population, thought to act as reserve stem cells of the intestinal and colonic crypts (192). Recently, high-complexity lentiviral barcodes relying on repeating sequences of degenerate bases have been

developed, increasing the resolution at which heterogeneity can be studied in these phenotypic clonal tracing studies (135,197). One design, originally developed for lineage tracing in the haematopoietic system (347) and later adapted for use in breast cancer PDX models (197), utilises 12 degenerate bases in a sequence 25 base pairs long, giving a theoretical complexity greater than 4 million. However, even this complexity limits the resolution at which clonal structures can be traced in a highly heterogeneous system.

In this chapter we derive and characterise tamoxifen resistant cell lines from an MCF7 parental line. We utilise the ClonTracer barcode library, featuring a 30-nucleotide degenerate-sequence and theoretical complexity of $\sim 72 \times 10^6$ (334), in order to track the contribution of phenotypically distinct cellular clones to the development of resistance. We reasoned that having multiple replicates with comparable starting barcode representations provides a means to distinguish pre-existing from *de novo* acquired-resistance clones. If resistance were mostly driven by *de novo* alterations, distinct barcoded populations would emerge in independent replicates; if, rather resistance was due to the outgrowth of a specific rare population, we would expect overlap in the enriched barcodes between replicates. Moreover, we use PyClone (335) and direct VAF measurements overtime to rule out genomic clonal selection, and putatively examine the contribution of quiescent, *Type IV* or slow cycling, cells to the resistance phenotype.

3.2 Characterisation of an Established Oestrogen-Insensitive Cell Line Derived from MCF7s

In order to establish a baseline for our studies, we sought to profile a previously described MCF7-derived tamoxifen resistant cell line, termed TamR (a generous gift from the Carroll Lab). Profiling would focus on ER α function, including cellular response to exogenous ligands (survival and proliferation) and downstream gene transcription of ER α target genes. TamR cells were cultured in charcoal-stripped FBS, phenol-red-free media as previously described (238), and as such it should be noted that they are likely to exhibit oestrogen independence³². In addition to resistance to tamoxifen, the TamR subline could be additionally resistant to, for example, AIs such as anastrozole, which work by inhibiting the generation of oestrogen from androgen precursors in the ovaries of premenopausal women and peripheral tissues of postmenopausal women (245).

Figure 3.2.1A depicts 4-OHT dose response studies, where charcoal-stripped media was swapped for growth media (containing FBS and phenol-red) and cells were treated with 4-OHT (afimoxifene) in escalating doses for 48hrs. The concentration at which 50% of cells were found to be viable (Inhibitory Concentration 50% (IC₅₀)) by trypan blue exclusion (348) was 0.24 μ M for the parental MCF7 cell line and 2.8 μ M for the TamR derivative ($p < 0.01$; extra sum-of-squares F test). Previous studies have placed the IC₅₀ for MCF7's as 0.39 μ M and several MCF7-derivative, tamoxifen resistant, lines in the 1.94 μ M to >10 μ M range (345), putting our results in keeping with established literature. ER α is known to bind multiple ligands simultaneously, reflected in the Hill Coefficient of >1 in each cell line. ER α has been shown, by competitive binding assay with oestradiol (E2) and 4-OHT at constant molar ratios, to preferentially bind 4-OHT, which competes with E2 for active binding domains on the receptor surface (349). Additionally,

³² For simplicity, and since the only agent tested was 4-OHT, we will refer to this (and cell lines derived in other studies) as tamoxifen resistant.

cooperative binding, where the binding of an initial ligand increases receptor affinity for the second, has been shown for both E2 and 4-OHT (349,350). Interestingly, the Hill Slope of ~4 in TamR cells treated with 4-OHT versus ~1.5 in MCF7s, suggests that this affect may be amplified in tamoxifen resistant cells.

To further investigate this phenomenon, and to continue characterisation of TamR and MCF7 cell lines for future experiments, we profiled *ESR1* (ER α) and *PGR* (PR) expression in each cell line, treated with combination or singular ligands, by qPCR. In each experiment, 4-OHT was used at 100nM (<50% of IC₅₀) and E2 at 10nM, based on studies estimating a typical biological concentration and showing differential gene regulation significantly above or below this level (351). Results in *Figure 3.2.1B* are an average of at least three independent biological replicates, and are presented as fold-change, relative to vehicle (EtOH) treatment of MCF7s, normalised to the geometric mean of *GAPDH* and *B2M* C_T values (both of which are commonly used housekeeper genes in breast cancer studies (352)). Treatment was for 24hrs, and statistical significance calculated as MCF7 versus TamR (One-way ANOVA with Bonferroni correction). Interestingly, *ESR1* expression was significantly, though not markedly, lower in TamR cells than MCF7s, at baseline ($p=0.0047$). Whereas E2 appeared to have relatively little effect on baseline expression of *ESR1* in either cell line, 4-OHT markedly reduced its expression in MCF7 cells only. A combination of E2 & 4-OHT similarly reduced expression of *ESR1* in MCF7s, though not to the same extent as 4-OHT alone, suggesting competitive antagonism between ligands. Overall, expression of *ESR1* expression was relatively unaltered by the presence or absence of either ligand in TamR cells. To test the function of ER α -related transcription, we next profiled cells under the same conditions, for *PGR* expression, as a classical marker of ER α function (353). As with *ESR1*, TamR has lower expression of *PGR* in baseline (EtOH) conditions ($p<0.00001$), the addition of E2, but not 4-OHT, increased *PGR* expression in MCF7s alone, as would be expected from functional ER α signalling. Interestingly, expression was reduced from baseline for TamRs treated with either ligand, alone or in combination, suggesting dysfunctional ER α signalling in this cell line.

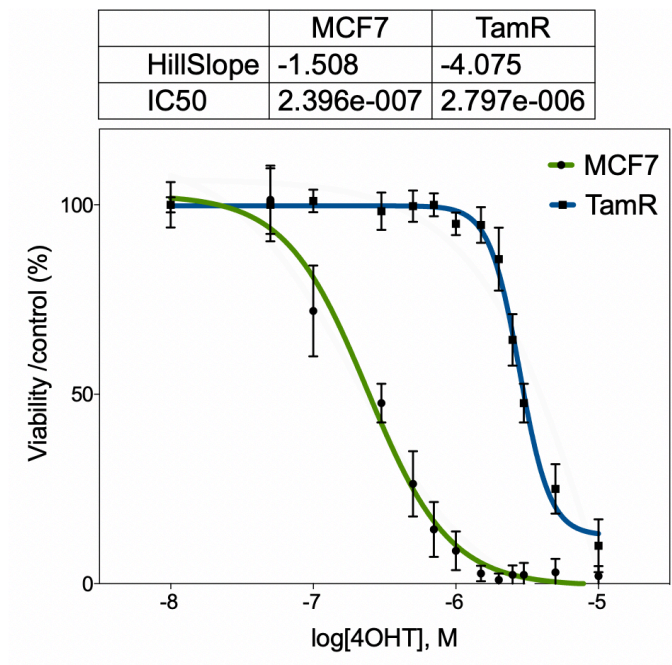
To test this hypothesis functionally, each cell line was cultured in the IncuCyte live imaging system for up to 150hrs in the presence of ligand, and assayed for proliferation (354). Whilst neither ligand appeared to have an effect on proliferation of TamRs, 4-OHT markedly slowed the growth of MCF7s ($p < 0.001$; non-linear mixed effects model) and exogenous E2 slightly increased their proliferation ($p > 0.05$). Together, these findings suggest the presence of a functionally different ER α signalling pathway in TamRs versus MCF7s. Whilst further studies, for example interrogating phosphorylation of downstream ER α targets, would aid us in a complete profile of the different signalling pathways, our aims are more to define a functional baseline with which to compare our own tamoxifen resistant MCF7 sub-lines.

Figure 3.2.1 – E2 & 4-OHT effects on MCF7s & TamRs

A) Dose Response of TamR and MCF7, showing resistance in the TamR line. $p < 0.01$; extra sum-of-squares F test. IC50 value is molar concentration. **B)** Expression by qPCR of ESR1 and PGR (ER α target) in MCF7s and TamRs treated with 10nM E2, 100nM 4-OHT or a combination. Shows loss of typical oestrogen response in TamRs. normalised to the geometric mean of GAPDH and B2M C_T values. Treatment was for 24hrs, and statistical significance calculated as MCF7 versus TamR (One-way ANOVA with Bonferroni post-hoc). **Figure Overleaf.**

Figure 3.2.1 – E2 & 4-OHT effects on MCF7s & TamRs

A



B

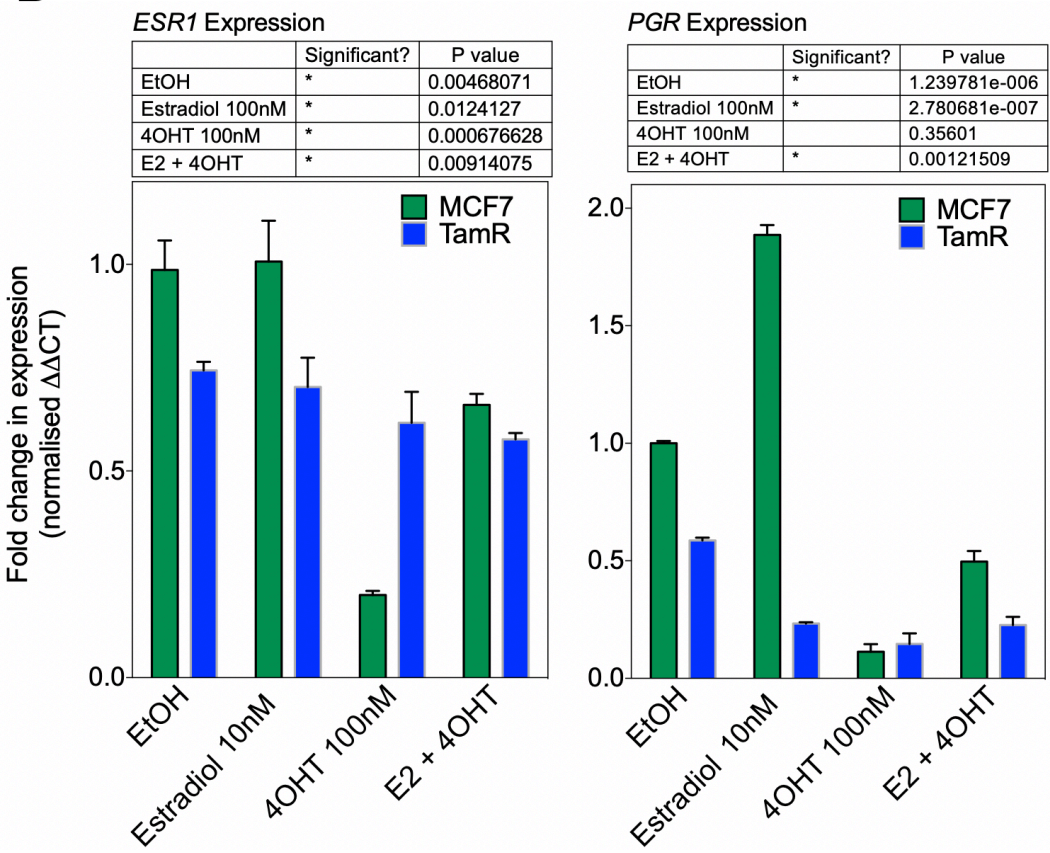
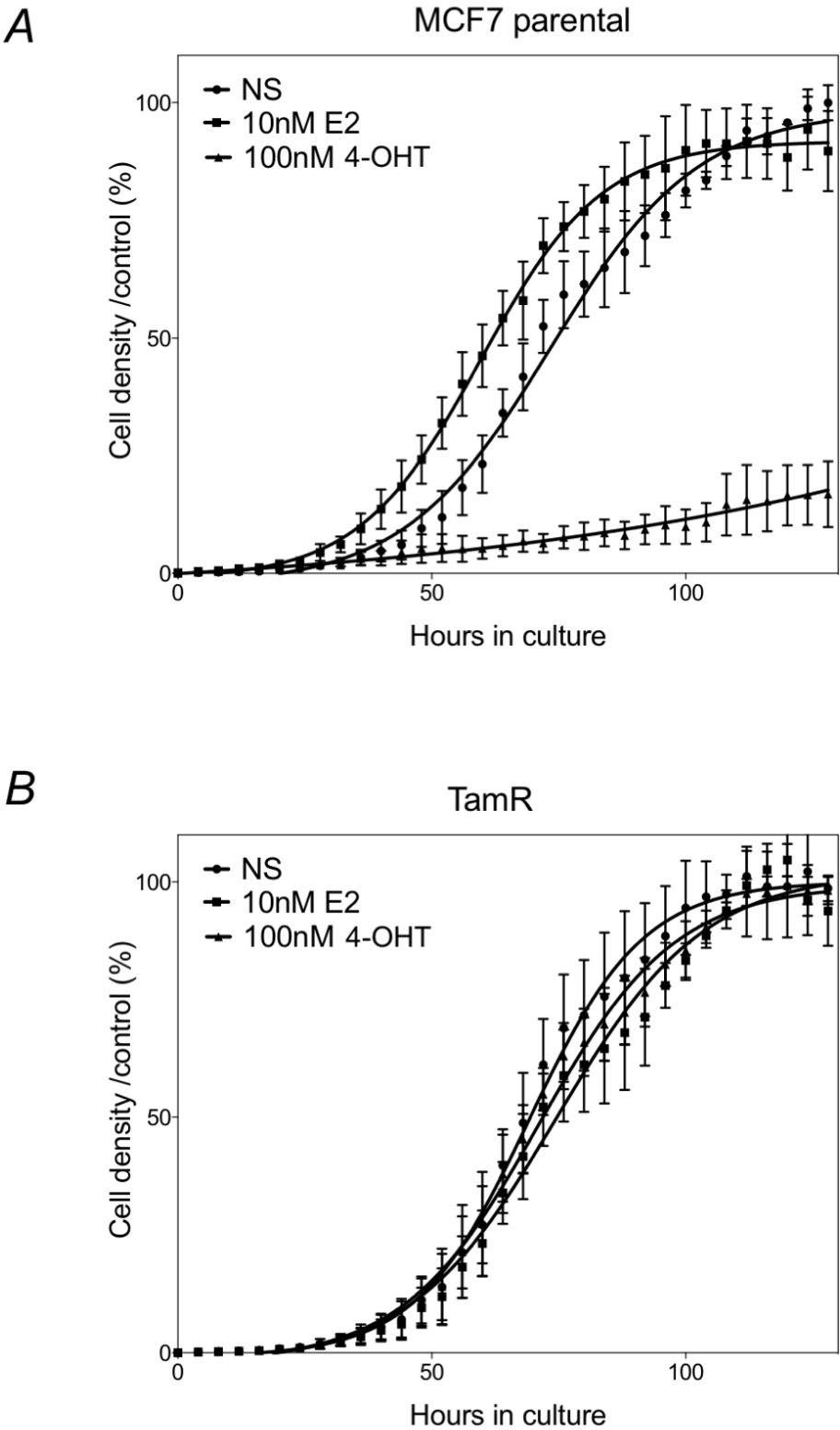


Figure 3.2.2 – E2 & 4-OHT effects on MCF7 & TamR proliferation

Incucyte growth curves of MCF7 parental and TamR subline in response to 10nM E2 or 100nM 4-OHT. **A)** Shows E2 induced growth and 4-OHT inhibition of growth in MCF7s. Inhibition by 4-OHT was statistically different from NS (control) and 10nM E2 ($p < 0.001$; non-linear mixed effects model, E2 increases in proliferation were not significant. **B)** shows the same experimental set-up for the TamR cell line. No differences were statistically significant. **Figure Overleaf.**

Figure 3.2.2 – E2 & 4-OHT effects on MCF7 & TamR proliferation



3.3 Selection of Tamoxifen Resistant MCF7 Sub-Lines

To derive tamoxifen resistant MCF7s, we cultured cells in T75 flasks for up to 160 days in the presence of charcoal stripped serum (c/s serum) and phenol red free media. Phenol red, a pH indicator normally included in cell culture media, has been shown to act as a weak oestrogen in cell culture (estrogenic activity at 15-45 μ M, affinity 0.001% of E2 in MCF7s) it is therefore recommended to conduct steroid hormone experiments in its absence (355). FBS contains multiple oestrogen and oestrogen-like compounds, we therefore opted for charcoal stripped serum, through this method is far from perfect, we sought to minimise the effects of remaining oestrogens by using serum from the same lot (356). In order to mimic more closely the physiological development of tamoxifen resistance in a breast cancer patient, we opted to include 4-OHT in our long-term resistance cultures. It was reasoned that even if c/s serum contained residual oestrogens, this would mimic the situation *in vivo*³³. 4-OHT was kept at a sub-IC50 dose and removed from the media 24hrs before any experiments to ensure residual drug did not to affect ER α binding experiments. *Figure 3.3.1* shows the culture protocol (A) alongside representative phase contrast images from defined time-points, and corresponding dose response graphs (B). In practice, cells were passaged after phase contrast images were taken, and re-plated to include experimental 24-well plates for parallel dose-response experiments, hence dose-responses were ~2-5 days after associated images were taken. Media was changed, and cells passaged in addition to the experimental procedure when required.

After 30 days culture in 4-OHT (*Figure 3.3.1A/B i*), MCF7s qualitatively showed slowed growth and pronounced budding, possibly indicating apoptosis. At this time-point, their IC50 for 4-OHT was practically unchanged from the control (MCF7s cultured in full

³³ As stated, the resultant tamoxifen resistant lines were likely resistant to both AIs and Tamoxifen, though due to the mechanism of action of AIs involving inhibition of oestrogen synthesis in the ovaries, testing *in vitro* would likely be inefficient, and as such these lines will be referred to simply as tamoxifen resistant.

serum & EtOH, as the 4-OHT carrier). At 60 days culture (*Figure 3.3.1A/B ii*) 4-OHT treated MCF7s showed increased cell growth, by qualitative analysis, and a slight increase in 4-OHT IC₅₀ (0.269 μ M to 0.397 μ M). Whilst far below the aforementioned IC₅₀ of TamR cell lines (2.8 μ M), this was an early indication of resistance, which at day 90 warranted an increase in the dose of 4-OHT from 400nM to 600nM, which was calculated to kill ~90% of parental MCF7s (IC₉₀). After 120 days culture (*Figure 3.3.1A/B iii*), clear cell growth was observed by phase contrast microscopy, separation in the dose response curves accompanied this pre-resistance phenotype (IC₅₀ = 0.67 μ M; $p < 0.001$ by extra sum-of-squares F-test). This was more pronounced after 160 days (*Figure 3.3.1A/B iv*) where the IC₅₀ for our cultured resistant cells was 1.876 μ M, and our control 0.311 μ M ($p < 0.001$), clear growth was also observed by phase contrast.

We sought to phenotypically profile the sub-cultured line (MCF7-R1) and to compare to established TamR lines. *Figure 3.3.2 A* shows the MCF7-R1 line cultured in the Incucyte cell imaging system, 'NS' refers to 'no-supplement' (EtOH carrier), results are relative to this EtOH control. EtOH controls were consistently no different from base line MCF7, data not shown. No difference was observed in proliferation of MCF7-R1 in the presence of either ligand. Expression of *ESR1* & *PGR* was profiled as before by qPCR, results are presented as normalised to internal housekeeper and EtOH control in *Figure 3.3.2B*. 4-OHT reduced expression of both genes, both on its own and in combination with E2. Whilst this was only seen in TamRs for *PGR*, there was a trend for decreased expression of *ESR1* also (*Figure 3.2.1*). It is possible that the discrepancy was caused by the comparatively prolonged (up to 40 passages post-induction) time that TamRs had been cultured in oestrogen-free conditions. Interestingly, whilst the Hill Slope found in each control (MCF7 parental) dose-response in *Figure 3.3.1* was close to 2 (1.98, 1.817, 2.135 & 1.8 for *i* through *iv*), the MCF7-R1 line progressed from 1.9 to 4.3 by the end of the resistance derivation period (though a dip below trend to from 3.6 to 2.1 was observed between 60 and 120 days), further reflecting the TamR line and hinting at increasing binding affinity in resistant cell lines.

Figure 3.3.1 – Deriving 4-OHT resistant MCF7s

A) Shows Phase Contrast of MCF7 over time with 4-OHT and a schematic for induction of resistance by continuous culture. Scale bar is 100 μ m. Time points are labelled i-iv, corresponding to dose-responses. IC50 values are molar concentration. **B)** Shows dose response curves for parental and continuously cultured sublines in response to 4-OHT treatment. Shows gradual induction of tamoxifen resistance. iii and iv (corresponding to 120- and 160-days culture) were statistically significant ($p < 0.001$, parental versus resistance by extra sum-of-squares F-test). **Figure Overleaf.**

Figure 3.3.1 – Deriving 4-OHT resistant MCF7s

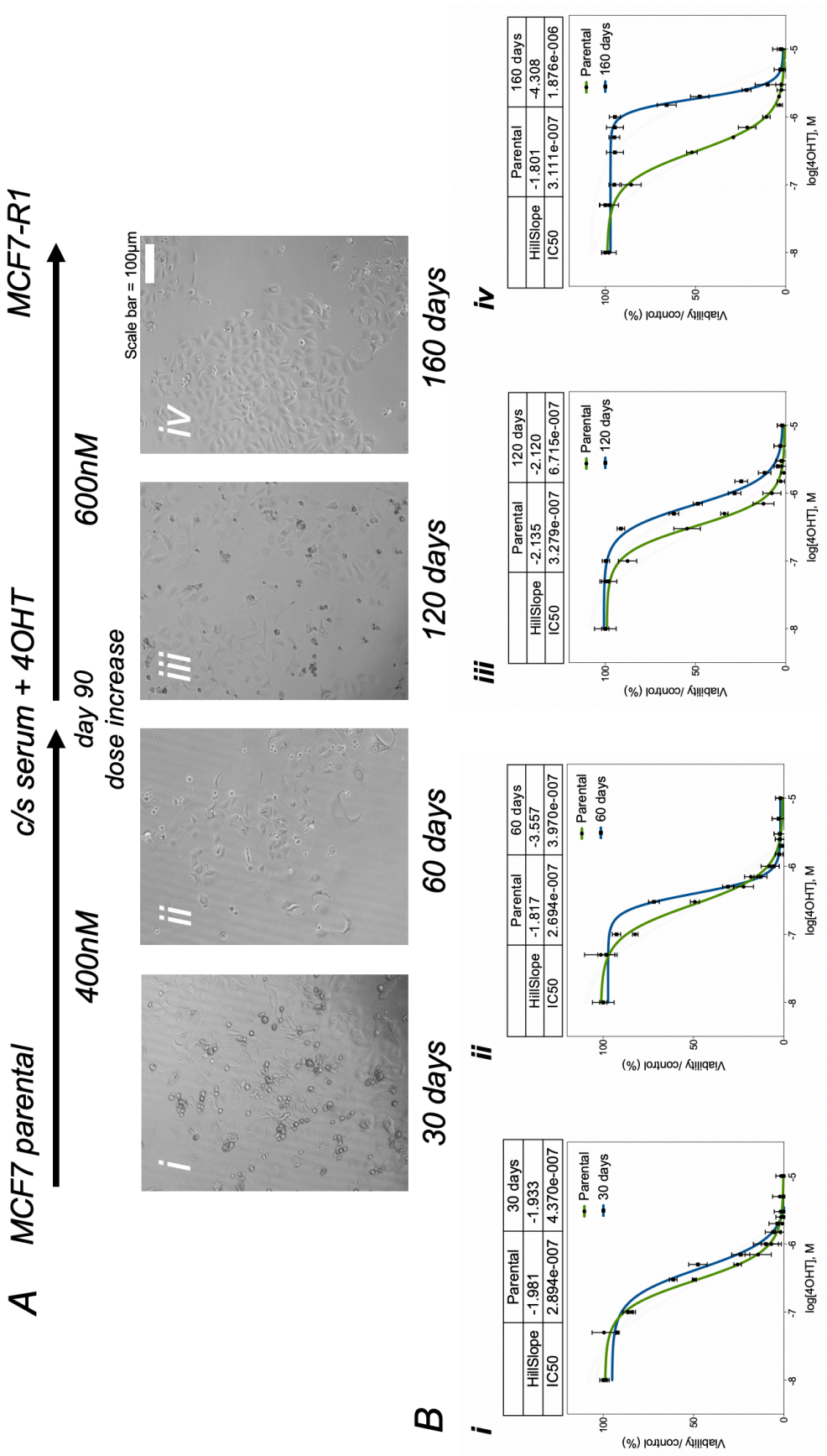
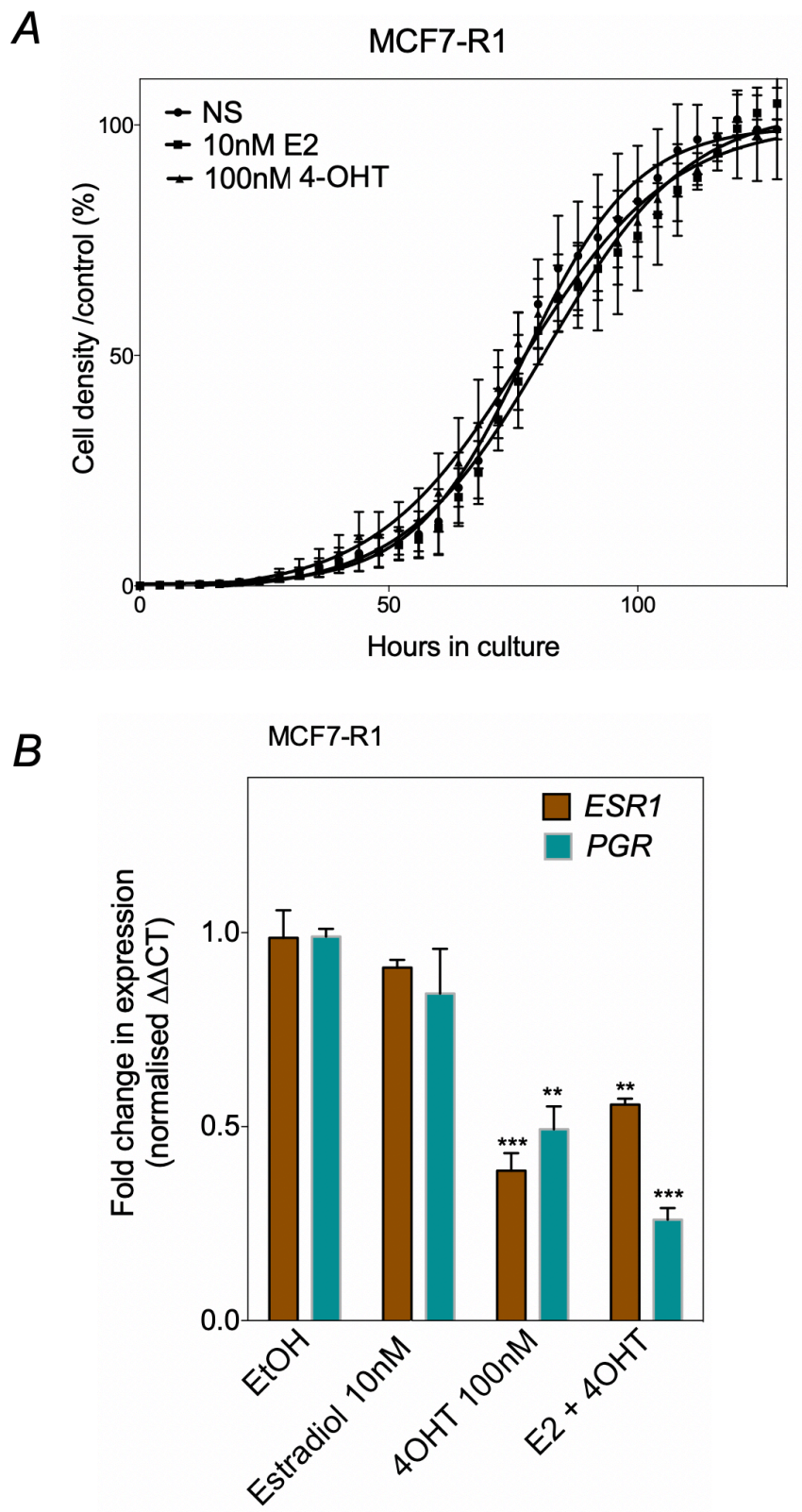


Figure 3.3.2 - Sensitivity of derived resistant lines to E2 & 4-OHT

A) Proliferation study with MCF7-R1, data derived from the IncuCyte live cell imaging system. No supplement (NS): i.e. EtOH carrier only. E2 dose was 10nM and 4-OHT 100nM, as before. No curves were statistically different by non-linear mixed effects model. **B)** qPCR RNA expression data confirming loss of normal oestrogen response in MCF7-R1 cells. Results are normalised to carrier (EtOH control) and GAPDH/B2M housekeeping controls and presented as fold change. 4-OHT significantly reduced expression of both ESR1 and PGR, alone or in combination with E2 (presented as versus EtOH control; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; one-way ANOVA with Bonferroni post-hoc). **Figure Overleaf.**

Figure 3.3.2 - Sensitivity of derived resistant lines to E2 & 4-OHT



3.4 Generation of ClonTracer MCF7 line (MCF7-CT)

Although targeted cancer therapies often yield impressive initial responses, tumours frequently develop resistance (357–359), this is no more obvious than in ER α + breast cancer, where the development of resistance to tamoxifen is common place (213,272). Typically, resistance results from the outgrowth of specific pre-existing populations within a tumour rather than from *de novo* evolution (135) (see *Figure 3.4.1 A*). Indeed, the wider the diversity of minor clonal populations in a tumour, the more likely it is that resistance will arise (11,124). However, the acquisition of *de novo* mutations, or indeed transcriptional programs, conferring drug resistance cannot be ruled out.

Tamoxifen resistant tumours are characterised by differential ER α -binding patterns with prognostic capacity exclusively in ER α + patients (248). Differential ER α -binding and the emergence of tamoxifen resistance in patients is not thought to be associated with the selection of a rare subpopulation of cells but rather rapid reprogramming of FOXA1 binding events (238). FOXA1 acts as a pioneer factor, recruiting ER α to the chromatin. However, what directs differential FOXA1 binding remains to be established (248). MCF7s have been observed to give rise to multiple distinct resistance phenotypes when cultured for prolonged periods with 4-OHT, supporting the idea that multiple tamoxifen resistant subpopulations exist within the parental cell line (345). Interestingly, MCF7s have an ‘intermediate’ ER α binding profile between ‘good prognosis’ and ‘bad prognosis’ profiles characterised by *Ross-Innes et al.*, (248) which could reconcile this supposed contradiction. Whether clonal outgrowth or the initiation of *de novo* resistance pathways, the development of drug resistance has profound implications for patients and warrants further study.

In this section, we utilise the ClonTracer barcode library to individually label a population of MCF7s (*Figure 3.4.1B & C*). In order to have enough starting material for lentiviral preparation and infection, we amplified the ClonTracer plasmid library as

described in *Methods (Chapter 2, section 2.3.1)* and ensured suitable library balance and complexity. 24 individual colonies were picked from across the amplification plates, plasmids purified and analysed by qPCR. *Figure 3.4.2 A* shows similar C_T values for each preparation, suggesting no significant bias in PCR amplification efficiency. To test the complexity of our amplified ClonTracer library, we serially diluted the pool to a theoretical concentration of 100 plasmid molecules. In the most dilute sample, 83 unique barcodes were found out of a theoretical 100, in the most concentrated, 100,836 out of 100,000. A linear correlation ($R^2=0.9989$; $p=0.0005$) was seen between number of predicted plasmid molecules and number of unique counted barcodes by NGS (*Figure 3.4.2 B*). It is likely that deviation from expected templates can be largely accounted for by instrument insensitivities and pipetting/ dilution inaccuracies.

Next, we prepared lentiviral constructs of our ClonTracer library in HEK293-T cells and optimised infection conditions in MCF7s. Lentiviral infection is thought to follow a Poisson distribution, in our studies, a MOI of 0.1 was targeted to ensure that most cells with an integration event (~10%) had a single integration event (360). Various methods exist to estimate virus titer, the two most common are direct p24 measurements by ELISA and qPCR for viral RNA. A p24 ELISA measures all p24 in the sample regardless of whether or not it is incorporated into a viral particle. Consequently, titres based on p24 quantification tend to be overestimates since they can include free p24 and defective viral particles. Direct measurement of lentiviral RNA is one alternative to direct p24 measurement. In this approach, viral RNA is first converted to cDNA and then quantified using qPCR primers targeting specific viral components such as long terminal repeats (LTRs), *gag*, *woodchuck hepatitis virus posttranscriptional regulatory element (WPRE)*, antibiotic resistance-genes, or the transgene itself. Similar to the p24 assay, titration via measurement of viral RNA can overestimate the amount of infectious virus due to the inclusion of potentially defective particles. Virion concentration can be 10- to 1000-fold higher by such methods than by functional assays (361). The ClonTracer backbone

contains a puromycin resistance gene to aid in selection of infected cells, thus we opted to calculate a functional titer directly in MCF7 cells.

We identified MCF7 48hr puromycin toxicity through titration to be between 1000ng/ml and 2000ng/ml (*Figure 3.4.3 A*), in line with values found in the literature (362), and so, opted for 2000ng/ml for 48hrs in selection of infected MCF7s. A viral titre of stock (2×10^9 TU/ml in HEK293-Ts) diluted 1 in 1,000 in PBS/serum was selected to give an MOI of <0.1 and thus the majority of infected cells to carry a single barcode integration. The number of unique barcode sequences provides a quantitative readout of the number of clones the population originated from, and the counts per unique barcode offered a measure of the relative abundance of each clone, although double integrations could potentially be removed by downstream computational analysis. Confirmation of viral titre was confirmed by flow cytometry analysis of infected MCF7s for RFP positivity. An RFP reporter is also present in the ClonTracer construct. *Figure 3.4.4 A* shows our gating strategy, first a wide side and forward scatter populations was selected to minimise cell debris, followed by a dead cell exclusion assay using propidium iodine, and doublet exclusion by both forward and side scatter discrimination. *Figure 3.4.4 B* shows the proportion of RFP positive (RFP+) cells with an escalating dose of ClonTracer viral prep. This analysis found 10.1% of cells were RFP+ with a viral load of 2×10^6 TU, showing reasonable concordance with the 7.6% of cells resistant to puromycin with the same viral load. Differences could be due to relative insensitivity of puromycin titration compared to flow cytometry. However, in giving the more conservative estimate of MOI, and thus less likely to suffer from false positive infections, puromycin selection was chosen for future studies ³⁴.

Our optimised infection protocol was used to infect 10 x 15cm dishes. Cells were puromycin selected for 48hrs and dead cells removed, around 20 million viral integration events were estimated over the pool. Dilutions were prepared, DNA was

³⁴ Though our data suggests this would underestimate true MOI.

extracted, and barcodes analysed by NGS as before. Good concordance was seen (*Figure 3.4.5 A*) between observed and expected unique barcodes ($R^2=0.986$; $p=0.0007$). In total 96% of the $\sim 1 \times 10^6$ barcodes sequenced, around 96% were unique, with $\sim 3\%$ having just two reads (*Figure 3.4.5 B*). These data suggest that the majority of barcodes are represented only once in our MCF7 pool.

Figure 3.4.1 - ClonTracer cellular clone tracking schematic

A) Model of the selection of resistant clones in a tumour. Dead cells are coloured in shades of black, other colours represent clonal populations. **B)** ClonTracer library consists of semi-random 30-bp-long DNA barcodes with 15 repeats of A or T ("W" for weak)–G or C ("S" for strong). **C)** depiction of a labelled tumour for clonal tracing, each cell, and its progeny, are uniquely labelled irrespective of genotype (denoted by colours). **Figure Overleaf.**

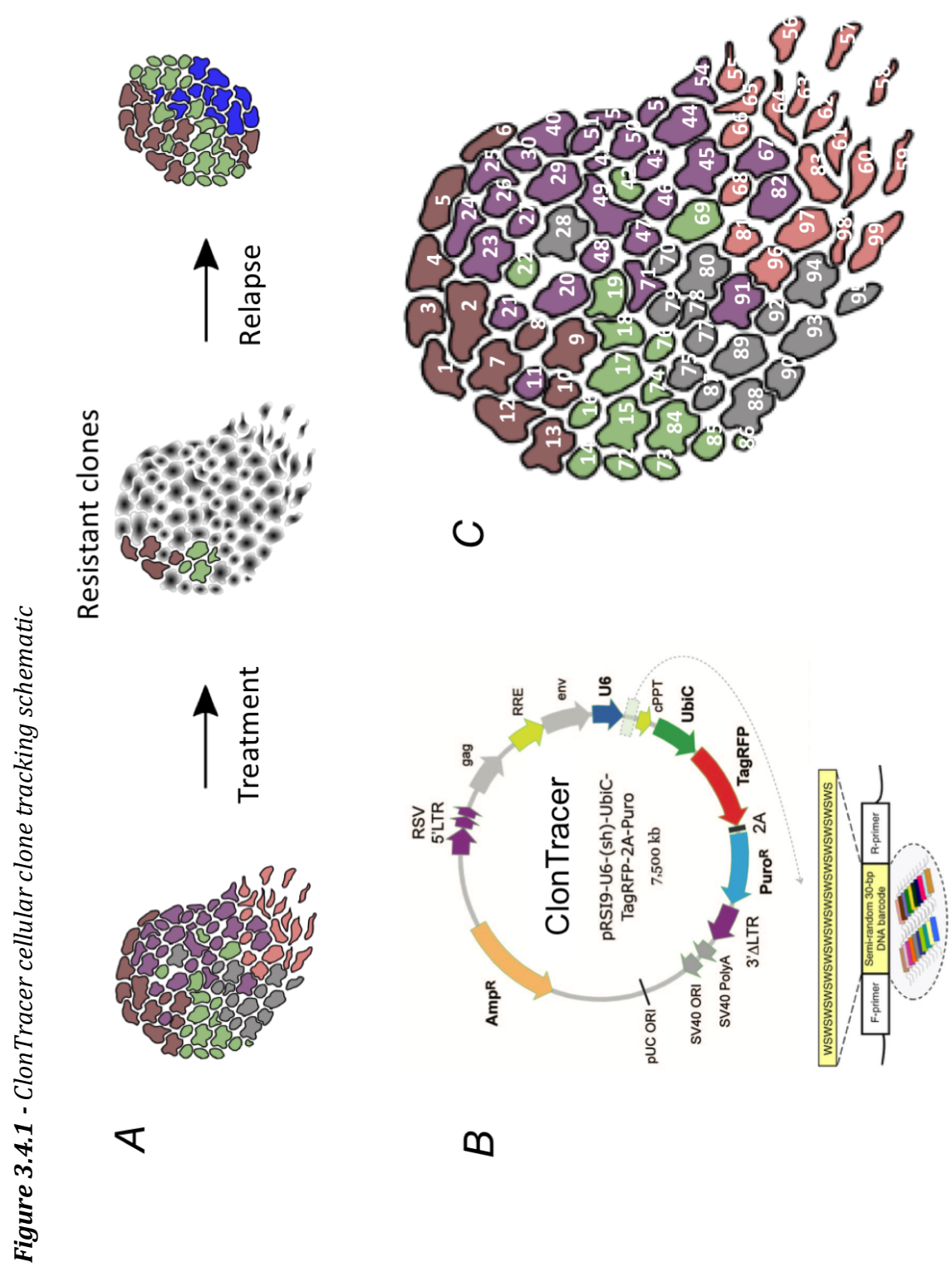


Figure 3.4.2 – Barcode amplification balance

A) 24 plasmid constructs containing a DNA barcode were isolated from *E. coli* colonies used in the expansion of the ClonTracer library. Remarkably similar CT values were observed for all 24 barcodes suggesting that the distinct barcode sequence did not introduce any significant bias in their PCR amplification efficiency. **B)** Plasmid stock was serially diluted to a calculated number of plasmids (by molecular weight). Plasmid dilutions were sequenced by NGS and unique barcodes counted using our barcode analysis pipeline (see: Methods, section 2.5). Good correlation between observed and expected unique plasmids ($R^2=0.9989$; $p=0.0005$; by linear regression) suggests the majority of barcodes are unique and library complexity is preserved through amplification protocols. **Figure Overleaf.**

Figure 3.4.2 – Barcode amplification balance

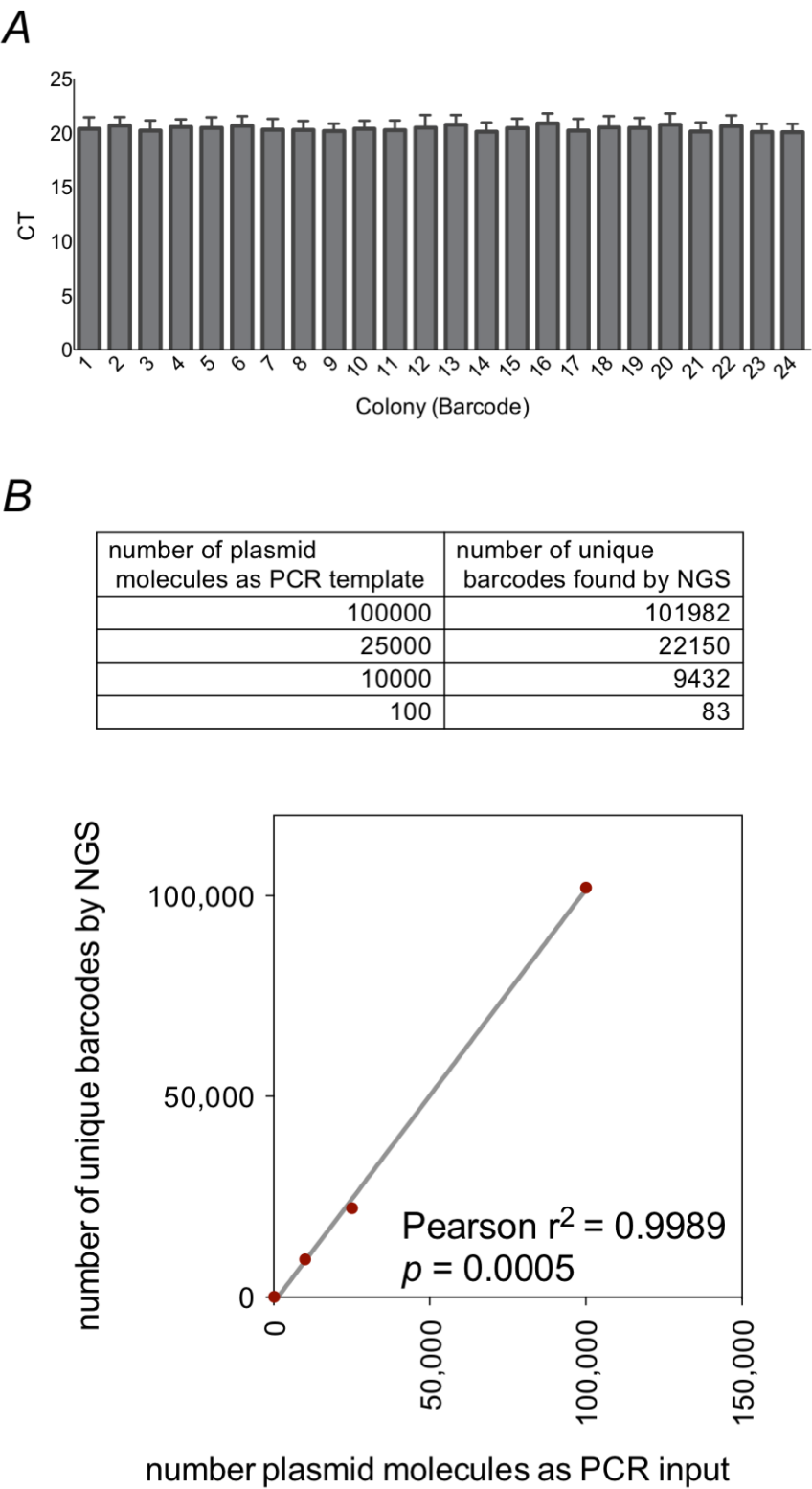
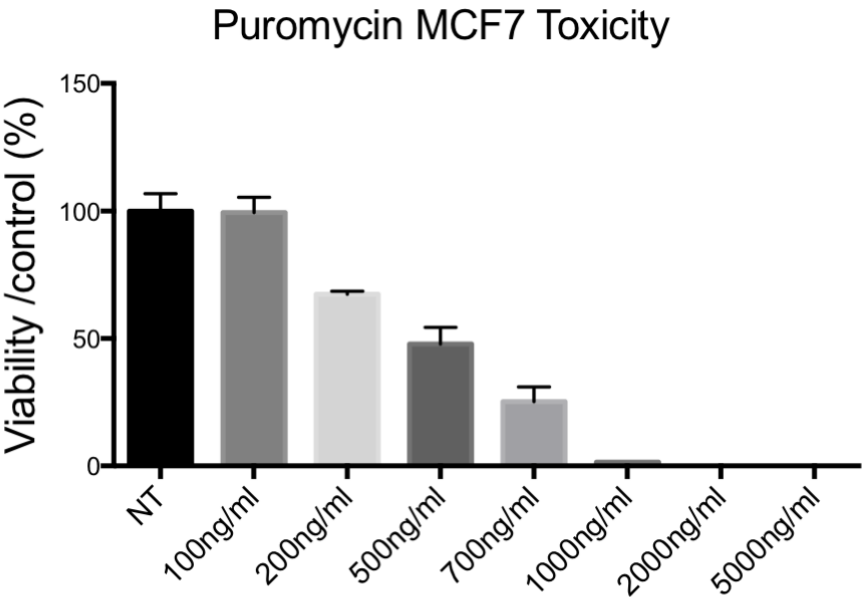


Figure 3.4.3 – Puromycin toxicity and viral titres for ClonTracer in MCF7

A) Shows puromycin toxicity for optimising infected MCF7 selection. The ClonTracer plasmid contains a puromycin resistance cassette to aid in both bacterial and mammalian cell selection. At 2000ng/ml <99% of MCF7s survived 48hrs of culture, as assessed by CellTitre-Glo. **B)** Viral titre of ClonTracer lentivirus, showing cut-off at MOI=0.1. Calculations (Methods, section 2.3.3) suggest this will result in the majority of cells being labelled by a single lentiviral integration. **Figure Overleaf.**

Figure 3.4.3 – Puromycin toxicity and viral titres for ClonTracer in MCF7

A



B

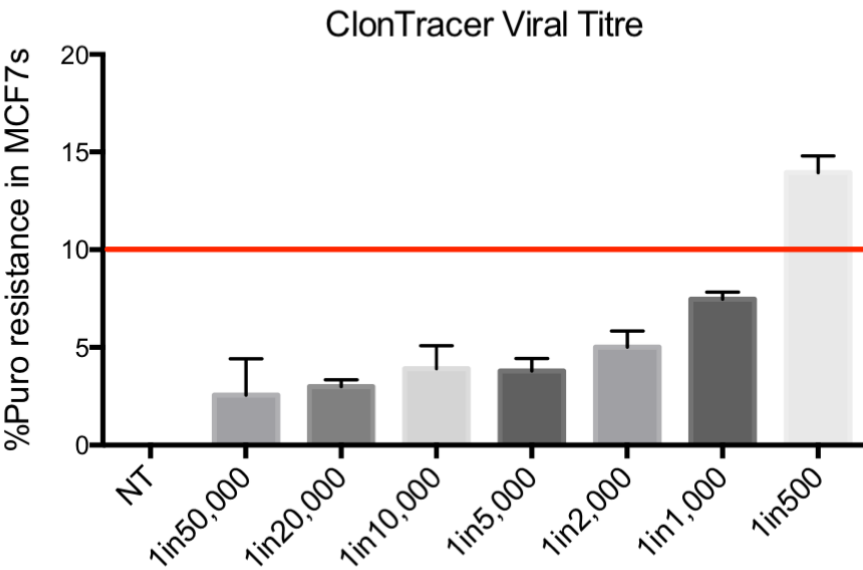


Figure 3.4.4 – ClonTracer MOI by flow cytometry

A) Shows our gating strategy, first a wide side and forward scatter population was selected to minimise cell debris, followed by a dead cell exclusion assay using propidium iodide, and doublet exclusion by both forward and side scatter discrimination. **B)** Shows the proportion of RFP+ cells with an escalating dose of ClonTracer viral preparation. Axis labels denote the wavelength of the excitation laser in nm. Excitation was collected at 588nm in both cases. **Figure Overleaf.**

Figure 3.4.4 – ClonTracer MOI by flow cytometry

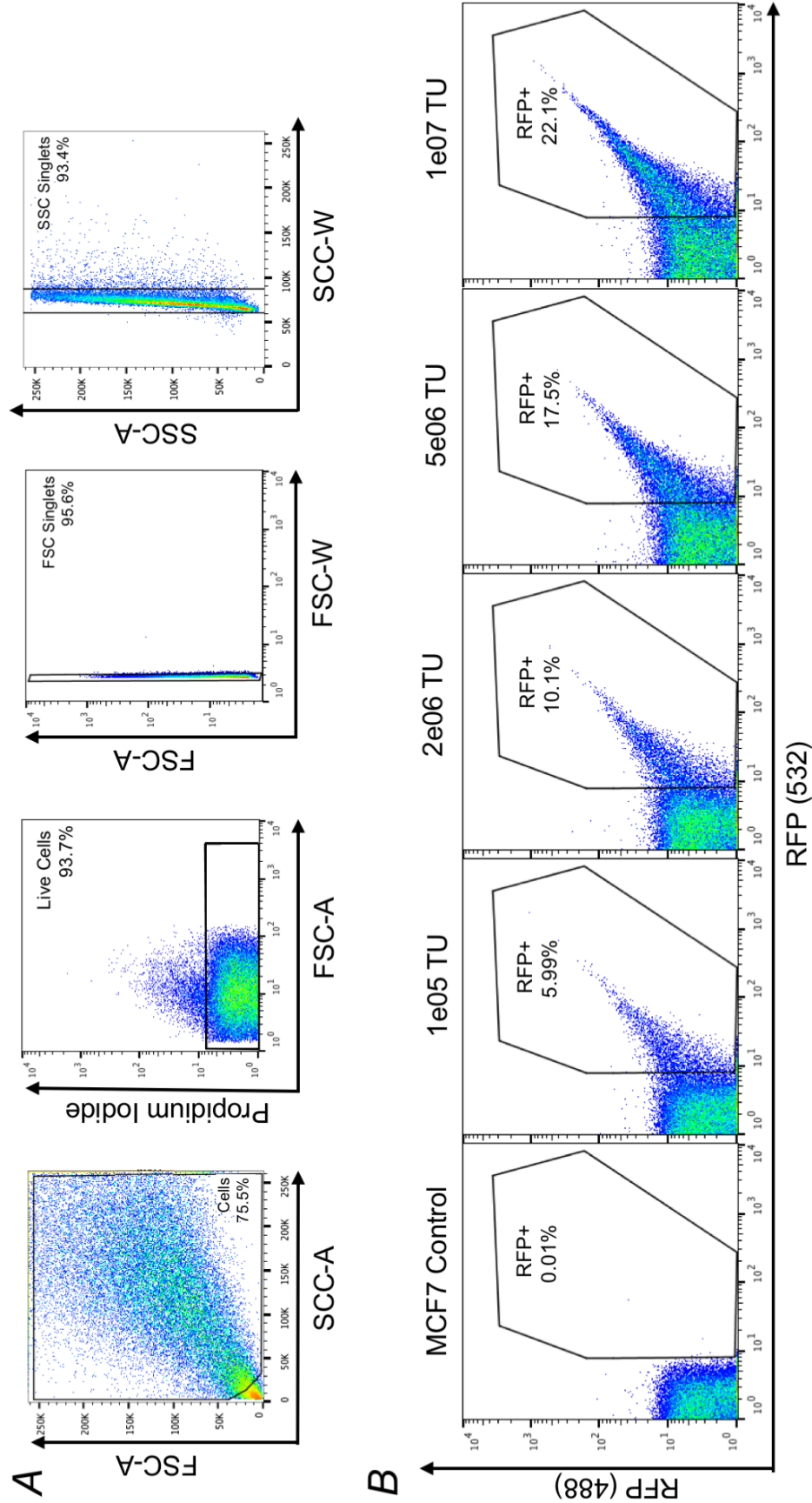
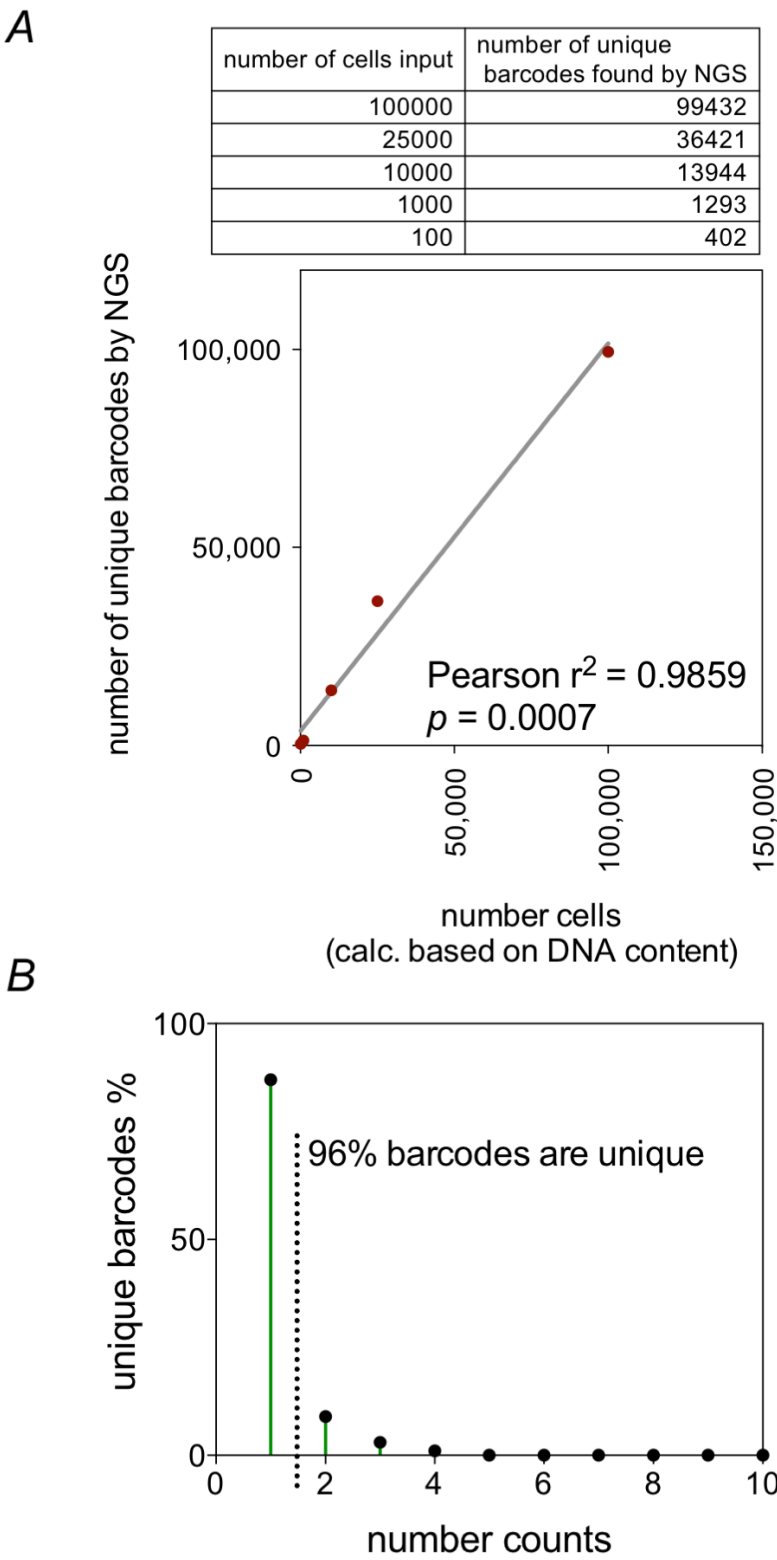


Figure 3.4.5 – Complexity of barcodes in MCF7 pool

A) Analysis as in 3.4.2 but for pool of 72hrs infected & puromycin-selected MCF7s (with dead cells removed & genomic DNA content used to estimate cell number). Good correlation is seen between observed and expected unique barcodes ($R^2=0.9859$; $p=0.0007$; by linear regression) suggesting that the majority of labelled cells are labelled uniquely. **B)** In total in this analysis, 4% of barcodes in the cell preparation were non-unique. Suggesting complexity is primarily maintained. **Figure Overleaf.**

Figure 3.4.5 – Complexity of barcodes in MCF7 pool



3.5 Tamoxifen Resistance Studies in Barcoded MCF7s

We sought to use the ClonTracer system to monitor clonal dynamics in response to 4-OHT treatment. We transduced MCF7 cells with the ClonTracer library with the aim of labelling approximately 10 million cells to deeply sample the potential heterogeneity of the cancer cell population, similar to the experimental plan designed by *Bhang et al.*, (334). We reasoned that having multiple replicates with comparable starting barcode representations would provide a means to distinguish pre-existing from *de novo* acquired-resistance clones. If resistance were mostly driven by *de novo* alterations, distinct barcoded populations would emerge in independent replicates. By contrast, if pre-existing clones were the major source of resistance, one could expect the selective enrichment of the same sets of barcodes (labelling the same pre-existing subpopulations) in replicate experiments.

To explore this, pooled barcoded MCF7s (MCF7-CT) were expanded and plated in 10 replicates with 10-fold library representation each (*i.e.* 10×10^7 cells each) to minimize stochastic loss of barcodes during plating and to ensure comparable starting barcode representations. *Figure 3.5.1* shows our experimental design. Plating was such that resistant and control replicates of the same number (*e.g.* MCF7-CT res1 and con1) were derived from the same originating Eppendorf, allowing direct comparisons to be drawn. To accommodate the need for 100 million cells per replicate, biological replicates (5 in each arm) consisted themselves of 5 x 15cm dishes (seeding density of 4×10^6). These technical replicates were maintained and expanded separately. Thus, 5x resistant and control pairs (consisting of 5x technical replicates) were established and propagated.

4-OHT resistance was established in the treatment arm as previously optimised, creating the “MCF7-CT res” cell lines (1-5), the control arm was treated exactly as treatment, but media (+serum) was spiked by EtOH rather than 4-OHT, giving the “MCF7-CT con” cell lines (1-5). Sampling by NGS was carried out after 3 and 6 months. *Figure 3.5.2* shows

the library preparation and bioinformatic pipelines for processing of barcodes. Aside from the FastX toolkit and standard NGS Quality Control (QC) checks, all processing was carried out in Python using the ClonTracer software package (135). Sample usable read counts are given for one NGS run in the bottom panel of *Figure 3.5.2*. Primer sequences, and more experimental details, are given in *Methods, section 2.5*.

After 6 months culture, we profiled each barcoded cell line (MCF7-CT res1-res5 & MCF7-CT con1-con5) for IC₅₀ by 4-OHT dose response. Each control-barcoded line was found to be sensitive to tamoxifen similarly to the parental line (*Figure 3.5.3 A*) and each resistant cell line showed an increased IC₅₀, suggesting barcodes did not interfere with the development of resistance (*Figure 3.5.3.2*). Resistant IC₅₀s for res1-res5 were 0.8 μ M ($p<0.005$), 1.5 μ M ($p<0.001$), 1.2 μ M ($p<0.001$), 1.6 μ M ($p<0.001$) and 1.1 μ M ($p<0.0005$); all statistical analysis by extra sum-of-squares F-test). Interestingly, the Hill Slope followed the same pattern as seen in *Figure 3.3.1* (increased slope /potential cooperative binding in resistant lines), with the exception of MCF7-CT res1, which also had the lowest calculated IC₅₀. These results are likely the result of less than perfect curve fitting, though could represent heterogeneity amongst resistance lines. Next, three representative resistance lines were grown to confluence in the Incucyte live cell imaging system (*Figure 3.5.4 A*). MCF7-CT res1-3 were slower growing than MCF7 parental line, though appear to be relatively similar in growth to TamR, MCF7 parental was significantly faster than any resistant lines ($p<0.05$; non-linear mixed effect model). The slow growth of cells cultured to tamoxifen resistance is a well-documented process (97) and could be linked to their observed adoption of a mesenchymal-like phenotype (319). Interestingly, MCF7-CT res2 was initially slower growing than any other resistant line, this is most noticeable between 42-54hrs of culture, where MCF7-CT res2 was >50% lower in confluence than the other resistant lines. This result was not statistically significant.

Control lines retained response to E2 and 4-OHT in terms of ER α function (measured by qPCR of *PGR* expression) (*Figure 3.5.4 B*) but this was lost in resistant lines. Specifically, MCF7-CT res1 showed slightly elevated *ESR1* and *PGR* expression on the addition of 4-OHT (or 4-OHT & E2), res2 showed the opposite pattern, with slightly lowered *ESR1* and *PGR* expression on 4-OHT treatment. Only downregulation of *ESR1* & *PGR* in 4-OHT & E2 treated MCF7-CT res2 was statistically different from the EtOH control (* $p < 0.05$ & ** $p < 0.01$ for *PGR* & *ESR1* respectively; one-way ANOVA with Bonferroni correction). In control lines, *PGR* expression was significantly elevated in both lines on E2 treatment, and significantly reduced on 4-OHT treatment and without E2). *ESR1* expression was reduced on 4-OHT treatment in both lines (with and without E2) but was only elevated by E2 treatment in MCF7-CT con1 (all statistics from one-way ANOVA with Bonferroni correction). In *Figure 3.5.4 C*, we see that resistant line MCF7-CT res4 has lost proliferative response to E2 and 4-OHT in culture, and these were retained by control line MCF7-CT con4. Specifically, E2 slightly increases proliferation of MCF7-CT con4 (* $p < 0.05$), where 4-OHT markedly reduces proliferation (** $p < 0.001$), neither of these effects were seen in MCF7-CT res4 (statistics by non-linear mixed effects model).

Reports in the literature have shown that derived tamoxifen resistant lines can differ markedly in sensitivity to rapamycin and activation of mTORC1/2. Moreover, Everolimus (rapamycin) is used clinically to tackle tamoxifen resistant breast cancers. Slight differences in resistant line phenotypes observed in *Figure 3.5.3.2* (IC50s of 4-OHT) and *Figure 3.5.4 A* (proliferation) led us to question whether the phenotypes of our five derived resistant cell lines were identical. *Figure 3.5.5* shows rapamycin dose response curves for 4-OHT resistant cell lines. Large differences exist between IC50s, which span from 0.8 μ M to 1.7 μ M. These results suggest that the mechanisms of resistance acquired by res1-res5 may not have been identical.

Figure 3.5.1 – Experimental design for resistance screen

*Schematic of screen for MCF7-CT lines. We aimed to label 10 million labelled cells. Puromycin selection was completed as previously described and resulting cells pooled. Plating was such that resistant and control replicates of the same number (e.g. MCF7-CT res1 and MCF7-CT con1) were derived from the same originating Eppendorf, allowing direct comparisons to be drawn. To accommodate the need for 100 million cells per replicate, biological replicates (5 in each arm) consisted themselves of 5 x 15cm dishes (seeding density of 4,000,000). These technical replicates were maintained and expanded separately. Thus, 5x resistant and control pairs (consisting of 5x technical replicates) were established and propagated. Resistant cells were maintained in culture for 6 months in charcoal stripped serum and with increasing concentrations of drug, during which time they became resistant to 4-OHT (MCF7-CT res lines). Control lines were cultured in serum containing media spiked with EtOH carrier. **Figure Overleaf.***

Figure 3.5.1 – Experimental design for resistance screen

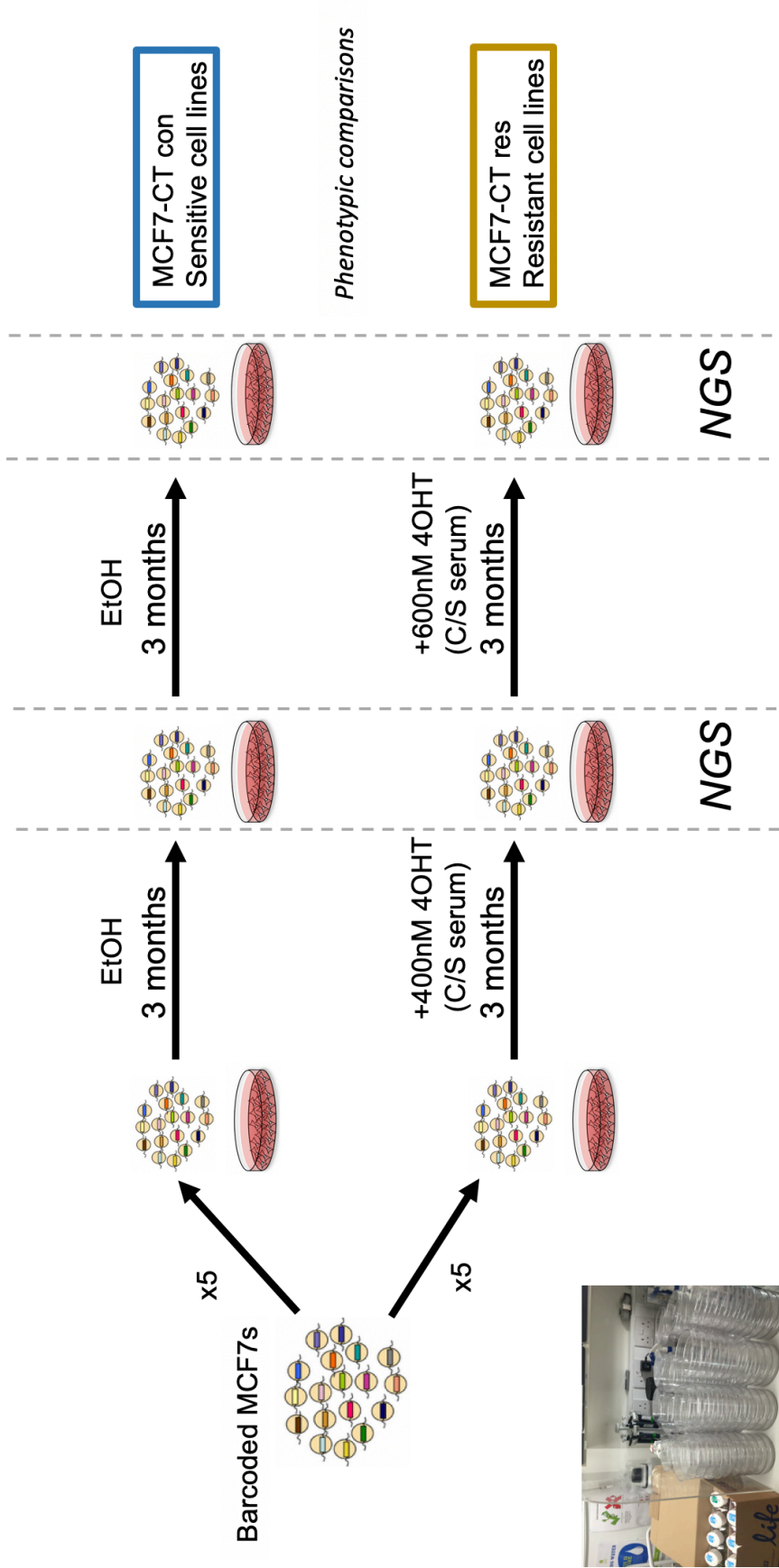


Figure 3.5.2 – Analysis pipeline for ClonTracer barcodes

Barcode of interest is amplified from genomic DNA by PCR, Illumina adapters and in-line barcodes are ligated before 150bp paired end (PE) sequencing. After standard QC and demultiplexing with FastX (barcode splitter function), total usable reads are tabulated, and barcodes extracted using the ClonTracer v1.2 python package. Counts and frequency distributions were then processed in R. Bottom panel shows representative output of one run, with usable reads (barcodes) for a range of replicates and experimental conditions.

Figure Overleaf.

Figure 3.5.2 – Analysis pipeline for ClonTracer barcodes

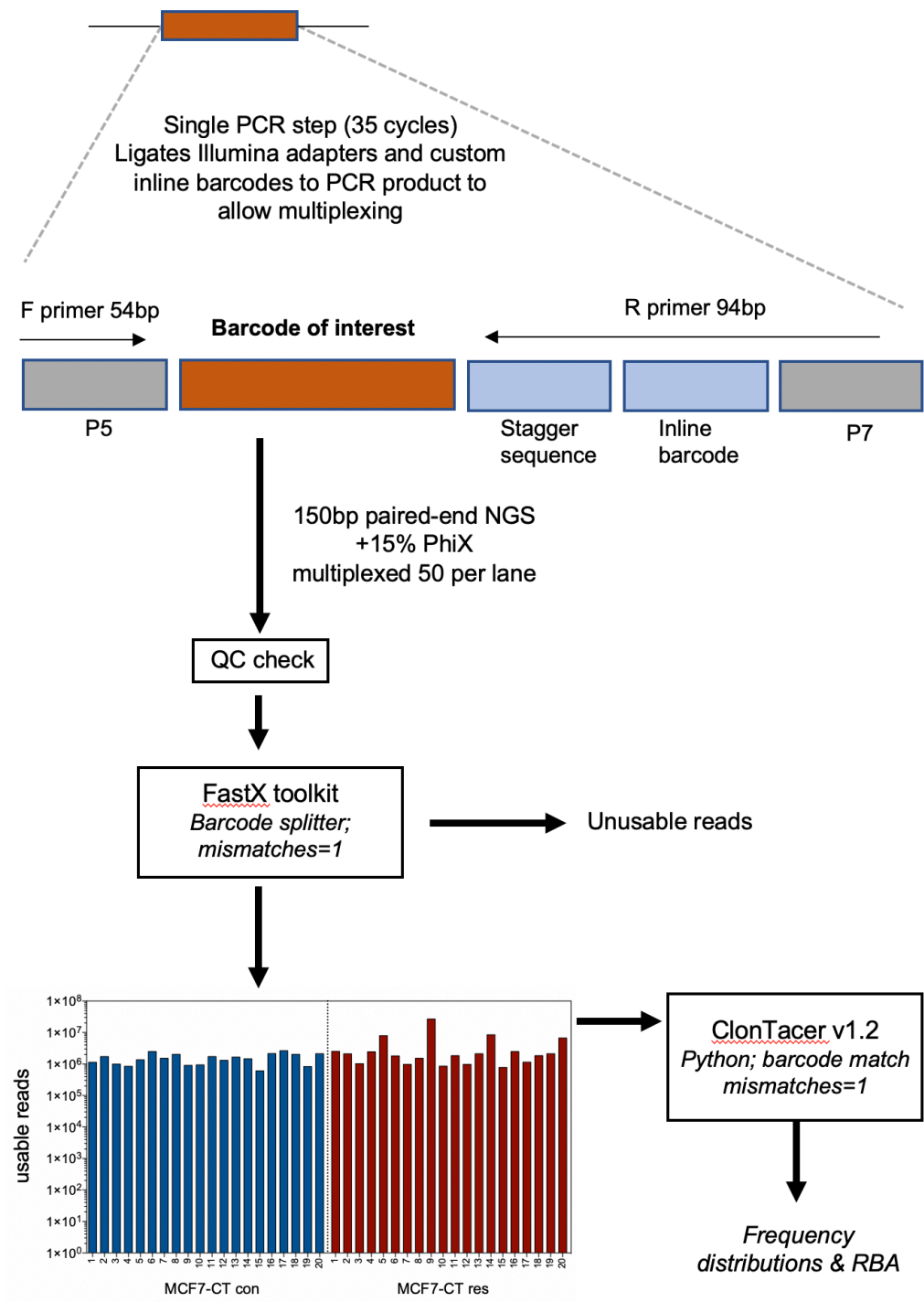


Figure 3.5.3.1 – 4-OHT response in sensitive lines, by dose response

Confirmation of 4-OHT sensitivity in MCF7-CT con cell lines. Dose response curves and calculations carried out as previously, in Figure 3.3.1. IC50 values are consistent with parental line in each comparison. These data suggest that prolonged culture of control cell lines has not affected their resistance to 4-OHT. **Figure Overleaf.**

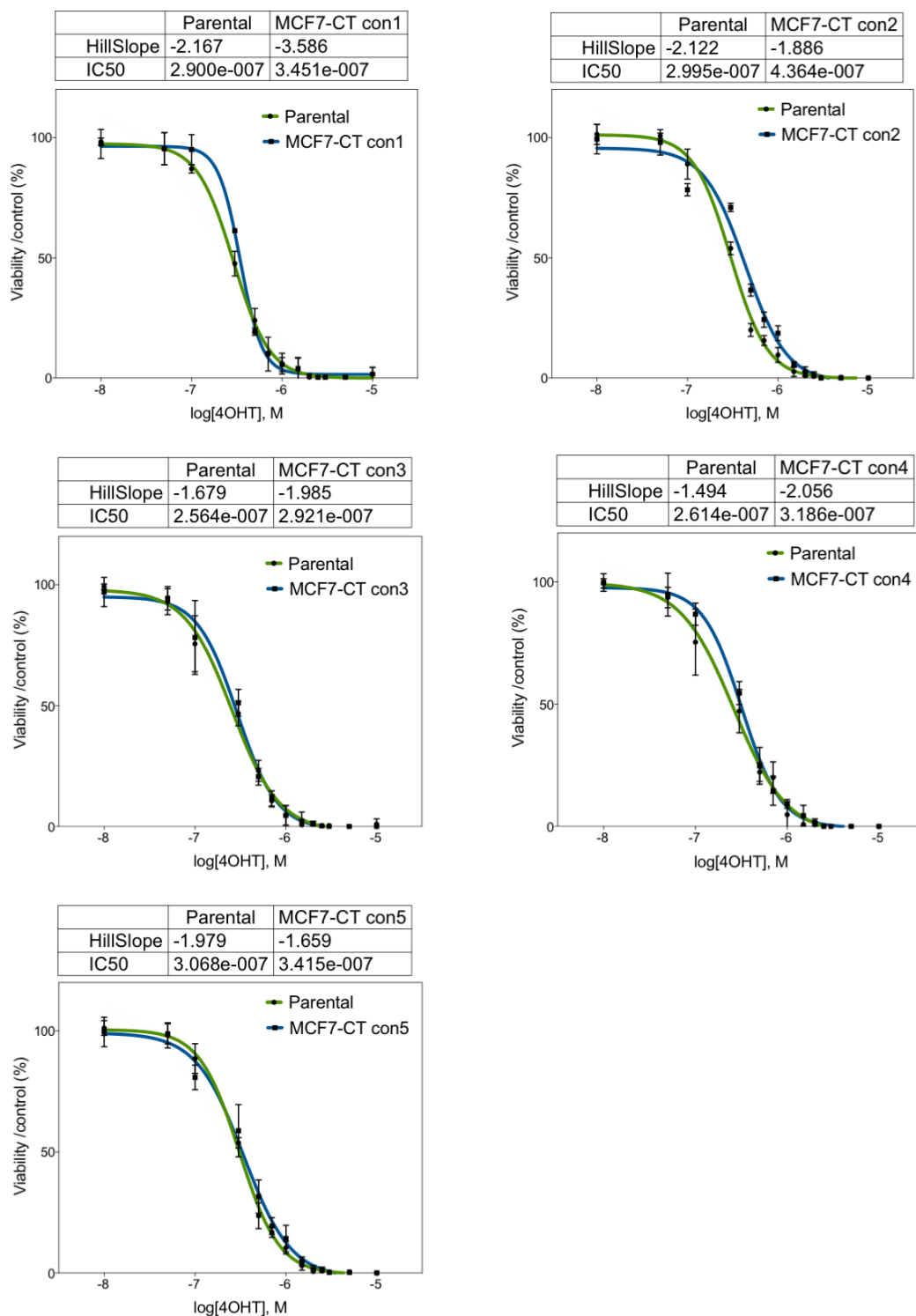
Figure 3.5.3.1 – 4-OHT response in sensitive lines, by dose response

Figure 3.5.3.2 – 4-OHT response in resistant lines, by dose response

*Confirmation of 4-OHT resistance in MCF7-CT res cell lines. Dose response curves and calculations carried out as previously Figure 3.3.1. IC₅₀ values are universally larger in MCF7-CT res cell lines, confirming the development of tamoxifen resistance in these lines. Specifically, IC₅₀s for res1-res5 were 0.8 μ M ($p < 0.005$), 1.5 μ M ($p < 0.001$), 1.2 μ M ($p < 0.001$), 1.6 μ M ($p < 0.001$) and 1.1 μ M ($p < 0.0005$); all statistical analysis by extra sum-of-squares F-test). **Figure Overleaf.***

Figure 3.5.3.2 – 4-OHT response in resistant lines, by dose response

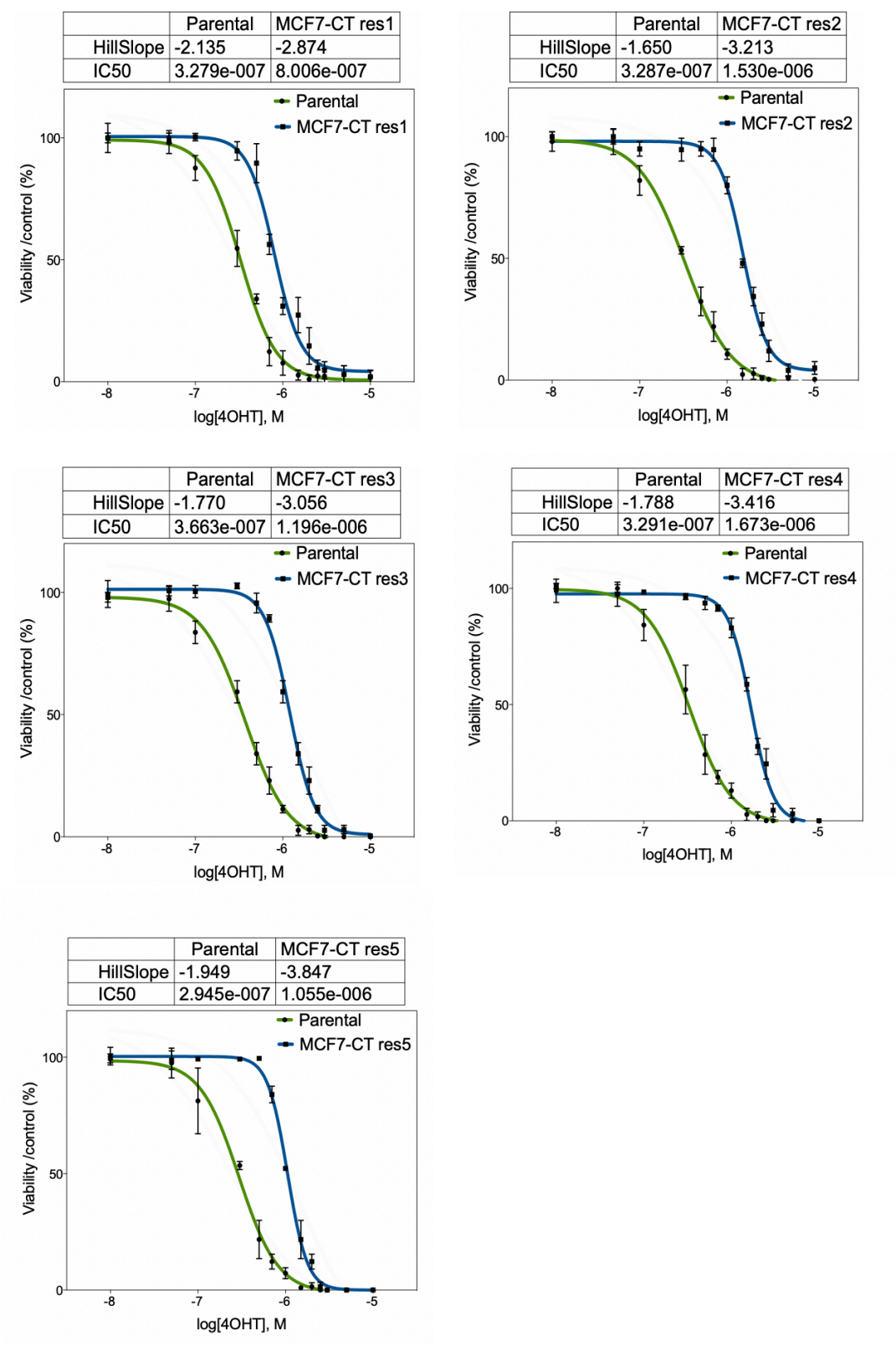


Figure 3.5.4 – Phenotypic response to estrogen of resistant lines

A) Proliferation (presented as relative confluence) of resistant lines (res1, res2 & res3), together with controls (MCF7 & TamR) over 90 hrs. MCF7-CT res1 & res3 appear similar in proliferative capacity to TamRs, whilst MCF7-CT res2 are initially slightly slower growing. Proliferation of MCF7s was significantly faster than all other cells lines ($p < 0.05$; non-linear mixed effect model), though the differences between resistance lines was not statistically significant in any comparison. **B)** Control lines retained response to E2 (10nM) and 4-OHT (100nM) in terms of ER α function. Res1 showed slightly elevated ESR1 and PGR expression with the addition of 4-OHT (or 4-OHT & E2), res2 showed the opposite pattern, with slightly lowered ESR1 and PGR on 4-OHT treatment. No resistance results were statistically significant. In control lines, PGR expression was significantly elevated in both lines on E2 treatment, and significantly reduced on 4-OHT treatment (with and without E2). ESR1 expression was reduced on 4-OHT treatment in both lines (with and without E2) but was only elevated by E2 treatment in con1 (all statistics from one-way ANOVA with Bonferroni correction). **C)** E2 slightly increases proliferation of MCF7-CT con4 ($p < 0.05$), where 4-OHT markedly reduces proliferation ($p < 0.001$), neither of these effects were seen in MCF7-CT res4 (statistics by non-linear mixed effects model). **Figure Overleaf.**

Figure 3.5.4 – Phenotypic response to estrogen of resistant lines

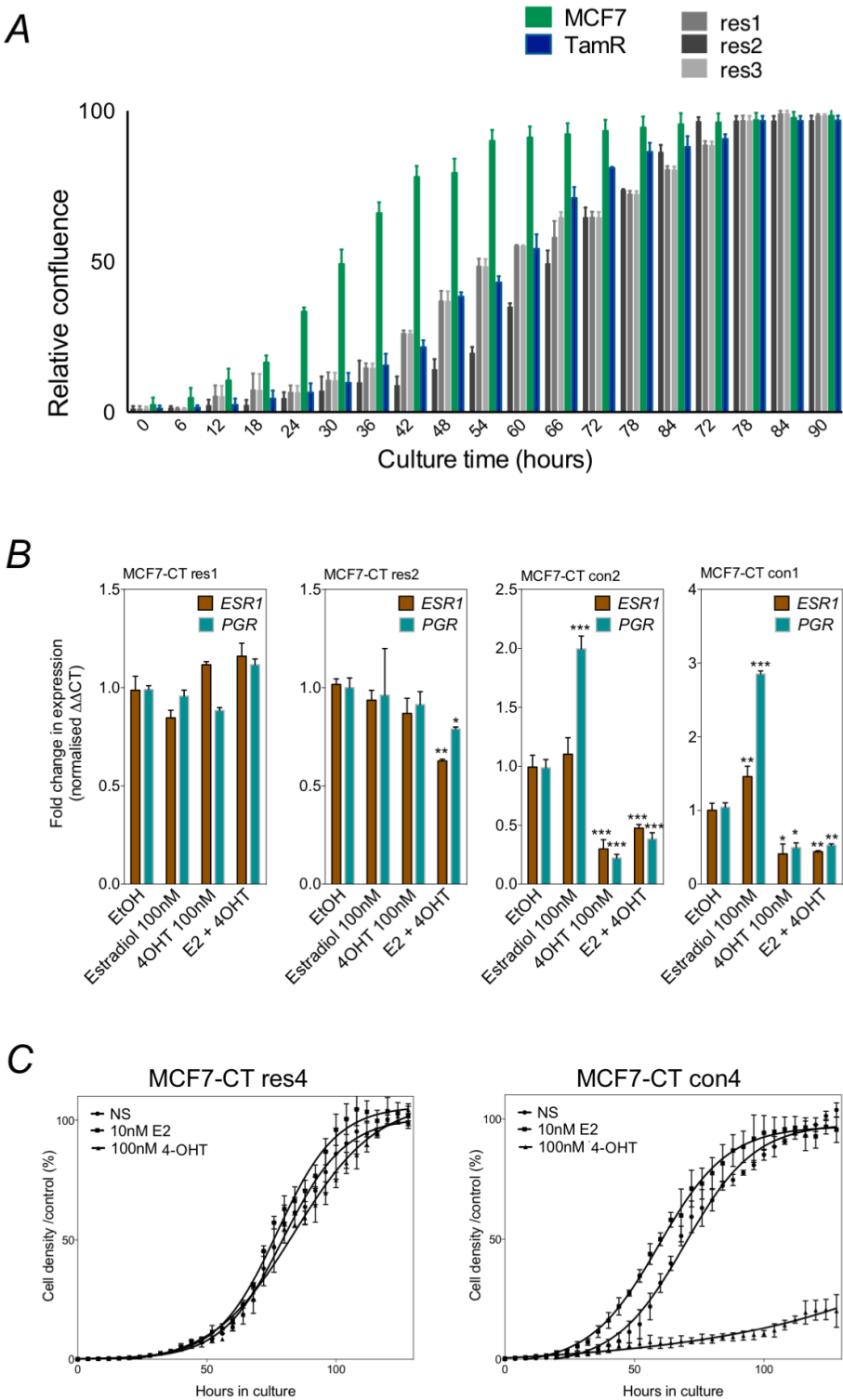


Figure 3.5.5 – Resistant lines differ in Rapamycin sensitivity

	TamR	Res1	Res2	Res3	Res4	Res5
HillSlope	-1.962	-2.488	-3.829	-1.899	-3.596	-1.973
IC50	5.808e-007	1.115e-006	1.750e-007	7.867e-007	1.674e-006	5.674e-007

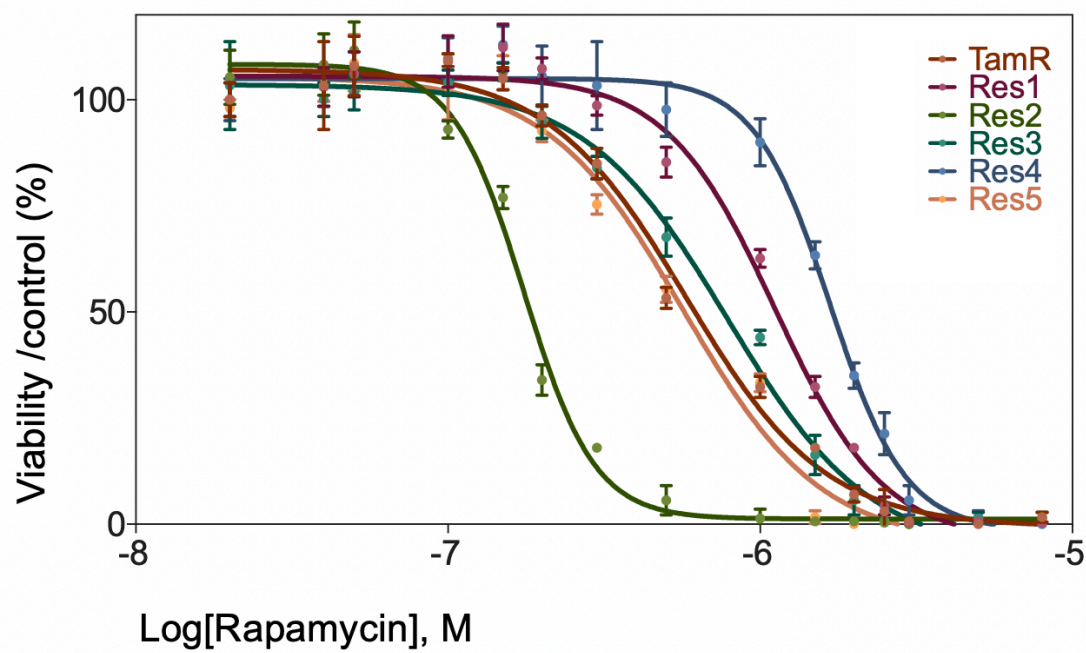


Figure 3.5.5 – Resistant lines differ in Rapamycin sensitivity

Rapamycin dose response curves for 4-OHT resistant cell lines. Large differences exist between IC50s, which span from 0.8 μ M to 1.7 μ M. Res2 has the lowest IC50 and Res4 the highest, comparison between these two curves was statistically significant by non-linear mixed effects model ($p < 0.01$) but this was lost on multiple comparison correction.

3.6 Barcode Distribution in 4-OHT Resistance

Following the development of tamoxifen resistance in our barcoded cell lines, we sought to quantify barcode abundance in each replicate. *Figure 3.6.1* shows relative barcode abundance (RBA), calculated as the proportion of total usable barcodes in the sample, represented by each unique barcode (along x-axis). RBA is a measure of the relative selection of an individual cellular clone (*i.e.* barcode) in a population, high abundance in few barcodes shows reduction of population complexity and, therefore, a selection for clonal populations. Technical replicates of MCF7-CT res1 (Rep1-Rep5) show selection of specific barcodes at abundancies ranging from 0.04 to 0.06, suggesting that 4-6% of cells in each replicate came from a single founding cell. Control lines (MCF7-CT con1, Rep1-5) show RBA consistently below 0.01, suggesting fewer than 1% cells originating from the same parental cell. These same data for all cell lines are presented in *Figure 3.6.2*, showing replicates 1-5 of each cell line in the same plot. Each line represents a specific unique barcode, with abundance represented by y-axis height. In each case, control lines show little specific barcode enrichment compared to resistant cells. Loss of complexity and a tendency toward clonal dominance can be modelled by the Shannon Diversity Index (SDI) (204,363). SDI has been computed for each cell line, with results printed at the top of each panel, lower number indicates a loss of diversity. Each control-resistance pair had a statistically significant reduction in barcode complexity and diversity: con1-res1 $p=0.045$; con1-res1 $p=0.009$; con1-res1 $p=0.007$; con1-res1 $p=0.008$ & con1-res1 $p=0.005$ (one-way ANOVA with Bonferroni correction).

Retroviral based therapies have been known to induce oncogenic transformation in a proportion of patients. It is thought that oncogene activation is caused by viral integration in specific genomic regions (364). Lentiviral based vectors are generally considered safer, they are known to integrate preferentially in megabase-wide chromosomal regions, and not in activating regions associated with cell growth (365). However, we sought to identify any potential cases of oncogenic viral integration in our

cell lines and exclude these from our analyses. *Figure 3.6.3 A* shows combined RBA data for control cell lines, we defined enrichment as any barcode present at a combined/cumulative RBA of 0.03. The 14 most enriched barcodes in the control population were therefor deemed to be enriched and removed from analyses. 10 of these 14 unique barcodes are also found (at varying RBA values) in the resistant cell lines, and so were excluded from downstream analyses (*Figure 3.6.3 B*). Clearly, the majority of barcodes enriched in the resistance cell lines were not also enriched in control lines, suggesting that the influence of any oncogene induced by lentiviral integration is negligible.

Figure 3.6.1 – Barcode enrichment in res1 & con1 cell lines after 6 months

Barcode counts (Relative Barcode Abundance; RBA) for MCF7-CT Res1 and Con1 were computed from NGS of 6-month cell cultures using the ClonTracer bioinformatics pipelines (Methods, section 2.5) (334). Relative barcode abundance is the proportion of a single unique barcode to the total usable barcode reads in a specific sample (this is a measure of clone size). **Figure Overleaf.**

Figure 3.6.1 – Barcode enrichment in *res1* & *con1* cell lines after 6 months

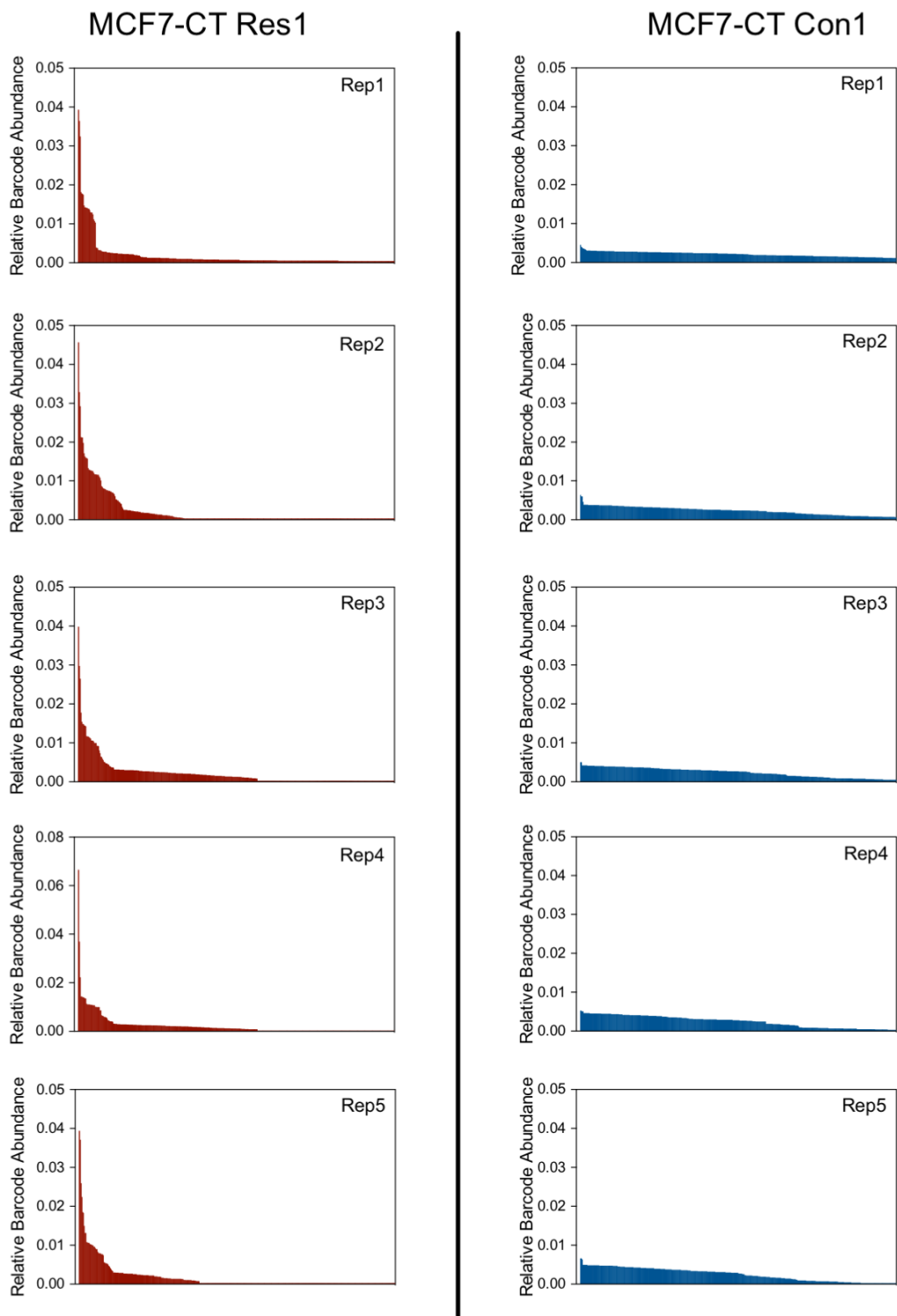


Figure 3.6.2 – Relative Barcode Abundance (RBA) in each cell line

*Distribution of RBAs across each cell lines and replicates. In each case, control lines show little specific barcode enrichment compared to resistant cells. Each line represents a specific unique barcode, with abundance represented by y-axis height. We computed the Shannon Diversity Index for each cell line as a measure of population diversity. The index is more often used in ecology studies but has been applied to cell line diversity studies previously (204,363). Enrichment of barcodes was compared by one-way ANOVA with Bonferroni correction to correct for multiple comparisons. Reduction of diversity is observed in each resistant line relative to its paired control line (con1-res1 $p=0.045$; con1-res1 $p=0.009$; con1-res1 $p=0.007$; con1-res1 $p=0.008$ & con1-res1 $p=0.005$). **Figure Overleaf.***

Figure 3.6.2 – Relative Barcode Abundance (RBA) in each cell line

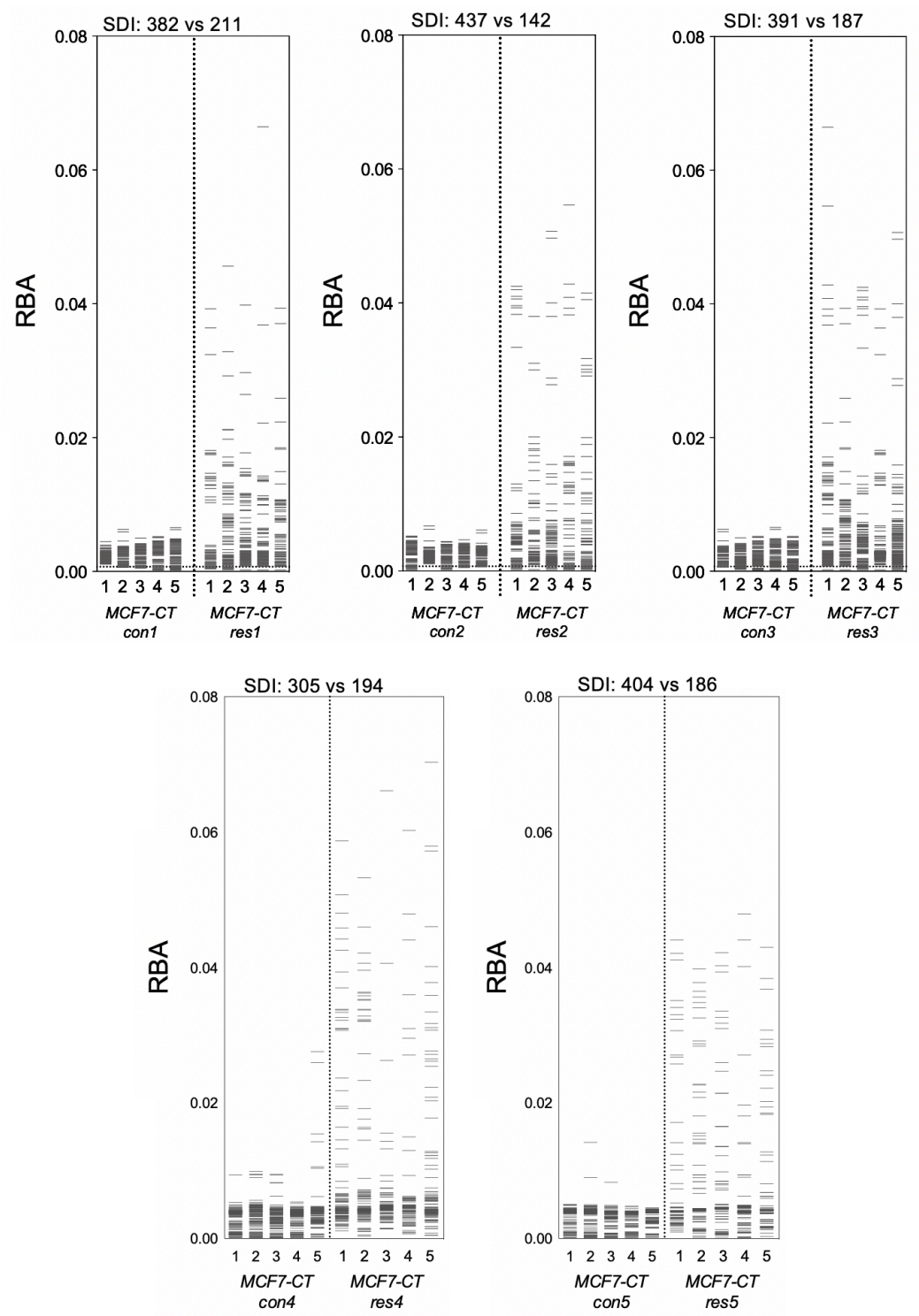
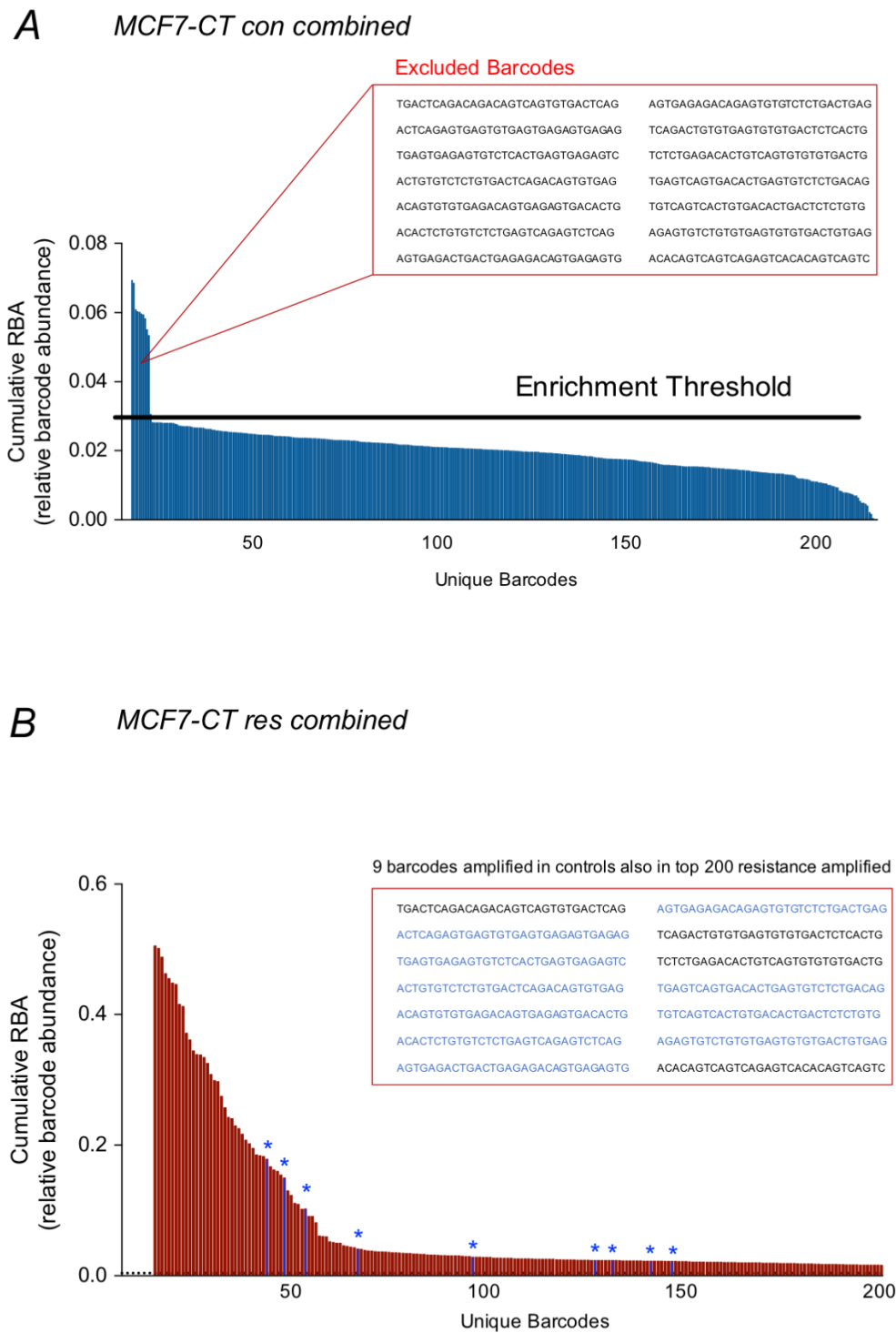


Figure 3.6.3 – Cumulative RBA across cell lines

*Cumulative enrichment of the top400 represented barcodes across independent replicates. RBAs were summed for each unique barcode across replicates. 14 barcodes are seen to be enriched in independent replicates of MCF7-CT controls and are excluded from further analysis. It is possible that PCR/analysis-based biases contributed to enrichment, though this is likely random. **Figure Overleaf.***

Figure 3.6.3 – Cumulative RBA across cell lines



3.7 Enrichment of Common Barcodes in Resistance

Although our tamoxifen resistant cell lines show substantial enrichment of specific barcodes relative to controls, we reasoned that this could have been due to the near-extinction events experienced by cells during the development of tamoxifen resistance. Because of the exponential nature of cellular growth, if a large proportion of cells in a population were killed or experienced slow growth for a prolonged period, population recovery could stochastically advance a sub population of cellular clones, and barcode abundance may not directly indicate cellular fitness. We reasoned that fitness would be indicated by barcode abundance should the same barcodes be enriched in multiple replicate plates. *Figure 3.7.1* shows the correlation between RBAs for the top 50 enriched barcodes from each of two replicates of MCF7-CT res1, two replicates from MCF7-CT res2, two replicates from MCF7-CT con1 & two replicates from MCF7-CT con2. Statistically significant association ($p < 0.0001$) was seen between each resistance replicate but not in control lines. Pearson's r scores were computed for each replicate (versus the other four) of each cell line and are presented in *Figure 3.7b*.

We expanded this analysis to all replicates of each cell line (*Figure 3.7.2.1* and *Figure 3.7.2.2*), this analysis was also expanded to the top400 barcodes (from top50 shown earlier). The mean correlation score (Pearson's r) between control replicates, *i.e.* the average intra-replicate concordance, was -0.01672, with the highest concordance of -0.0084 seen in MCF7-CT con5 (*Figure 3.7.2.1*), no control line showed statistically significant overlap in enriched barcodes. By contrast, overlap across the independent resistance replicates was highly statistically significant on the basis of a Pearson correlation test (mean $r = 0.41776$, $p < 0.00001$; *Figure 3.7.2.2*). The most concordant cell line was res3 (mean $r = 0.4824$, $p < 0.0001$) and least res2 (mean $r = 0.3756$, $p < 0.001$; *Figure 3.7b ii*). These results show a high degree of concordance between cells enriched in each replicate of the same cell line. We next calculated the proportion of the top100 enriched barcodes shared across all 5 resistant and all 5 control cell lines. *Figure 3.7.3 A*

shows a high proportion of the most enriched barcodes are shared between cell lines in the resistance models, but not the controls (*Figure 3.7.3 B*). Specifically, an average of 42% of the most frequent barcodes in one cell line were also enriched in each of the other four. Only 8% of the most frequent barcodes were uniquely enriched in that cell line only. In resistant lines, an average of 86% of barcodes in the top100 in one cell line were also in the top100 of at least one more cell line (*Figure 3.7.3 A*). <1% of barcodes were similarly enriched across more than a single control line (*Figure 3.7.3 B*). On average 49.6 top100 barcodes were enriched in all five technical replicates of a single resistance cell line, of these, 83% were enriched across more than one cell line (*Figure 3.7.3 C*). Together these results indicate a reproducible enrichment of discrete numbers of barcodes (*i.e.* cellular clones) in response to prolonged treatment with 4-OHT. As the same barcodes tend to be enriched across independent replicates, we can conclude that certain cells have pre-existing resistance, or are at least primed to become resistant.

Figure 3.7.1 – Correlation of barcode abundances between replicates

*Example correlations between RBA of the top 50 (by cumulative RBA across replicates) barcodes found in two sensitive and two resistant cell lines. Good Pearson's r score suggests shared abundance between replicates and deterministic selection in tamoxifen resistant lines only. **Figure Overleaf.***

Figure 3.7.1 – Correlation of barcode abundances between replicates

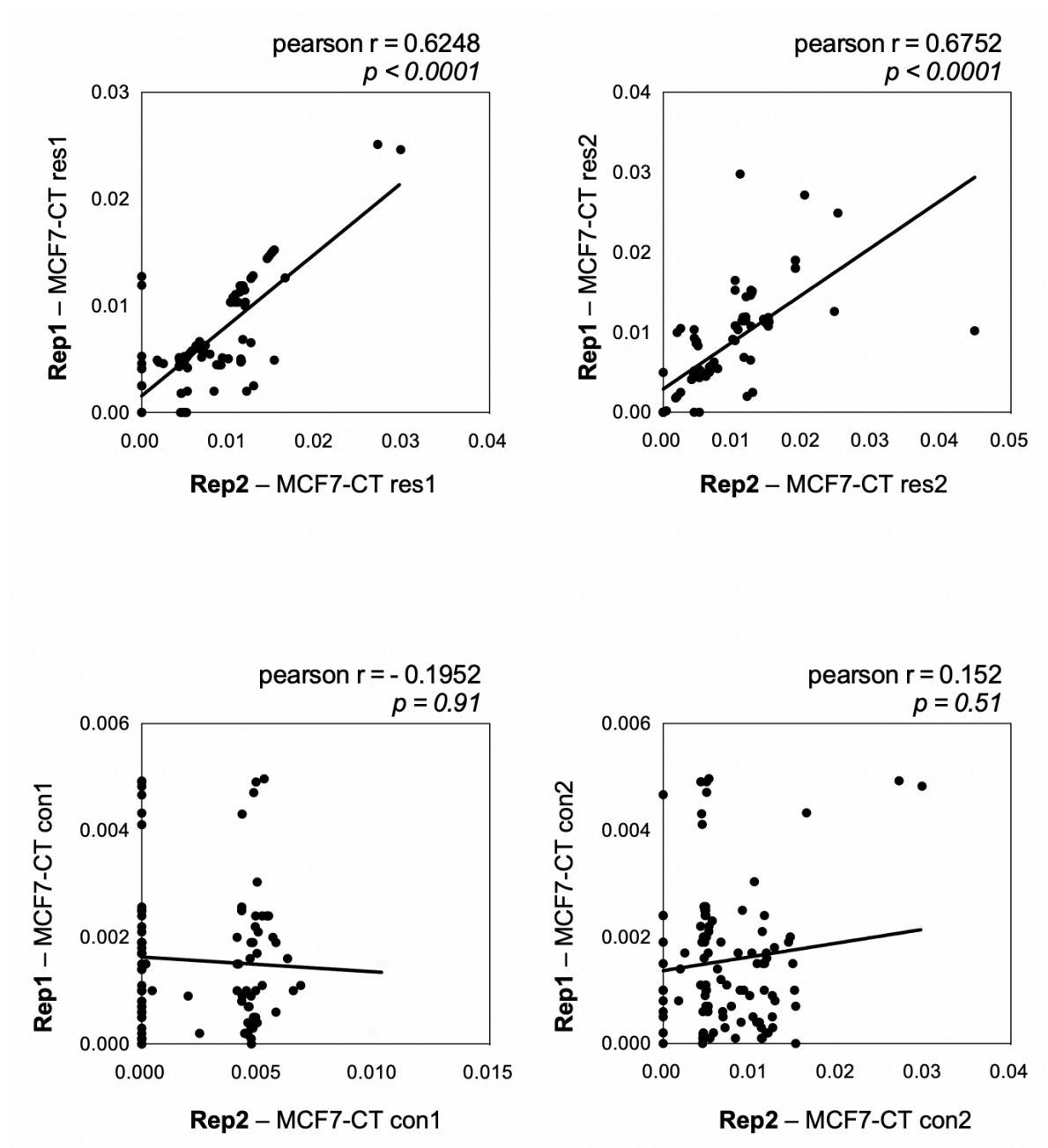
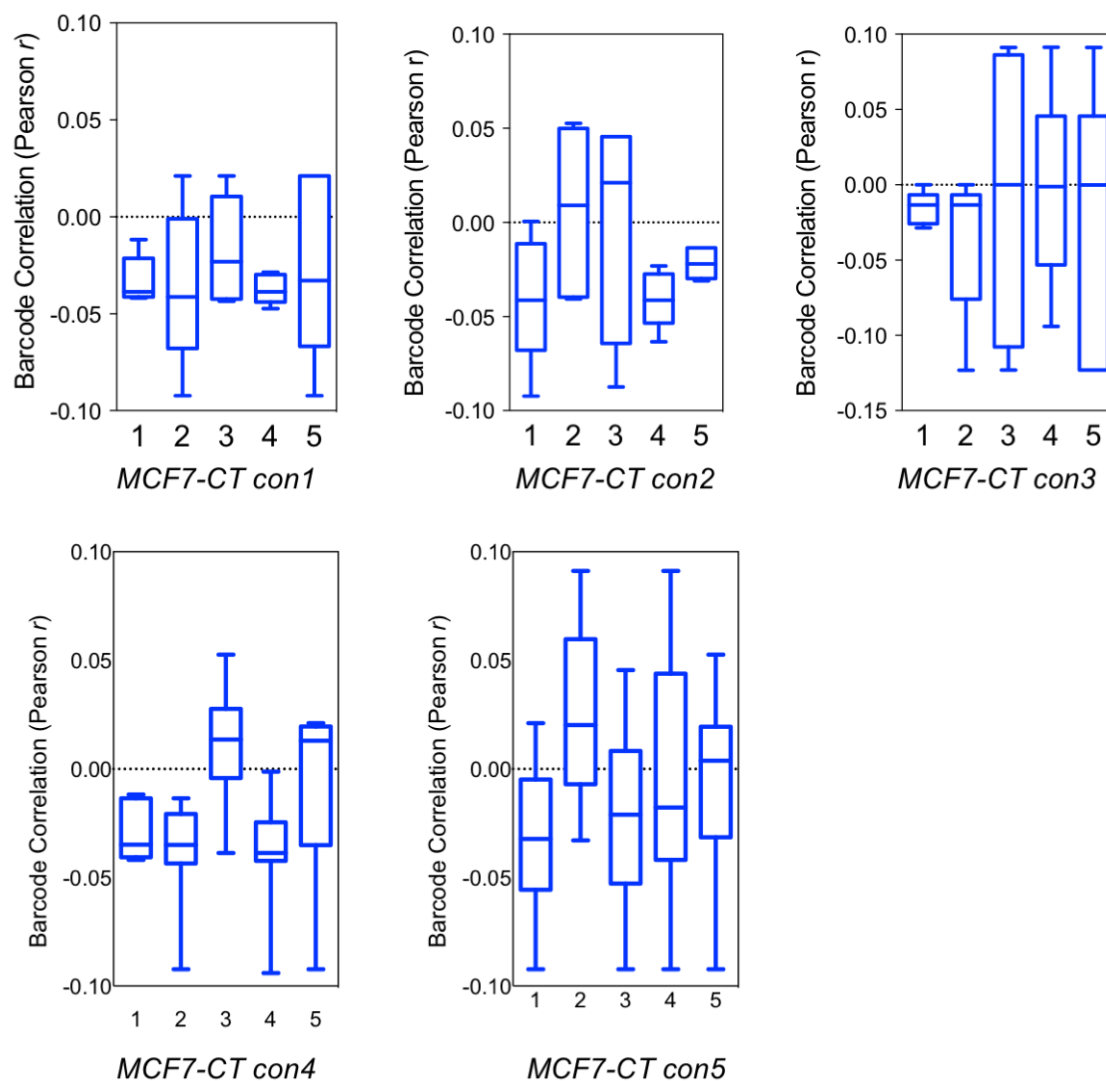
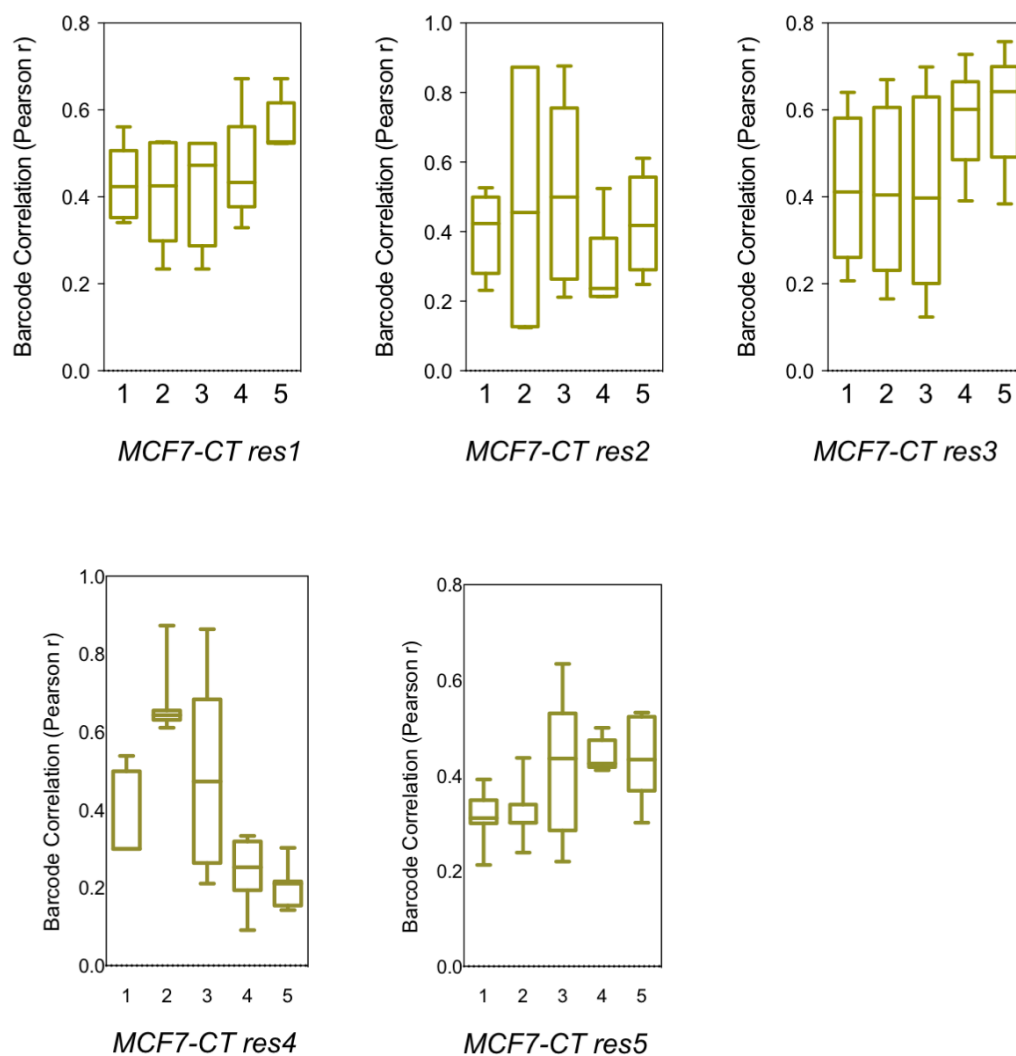


Figure 3.7.2.1 – Shared barcodes in sensitive cell line replicates**Figure 3.7.2.1** – Shared barcodes in sensitive cell line replicates

Combining Pearson's r scores for each pairwise comparison between control cell line replicates to create a boxplot. Results show the average correlation between each replicate and the other four replicates. Control lines show consistently low correlation between replicates, suggesting different clonal populations are enriched in each.

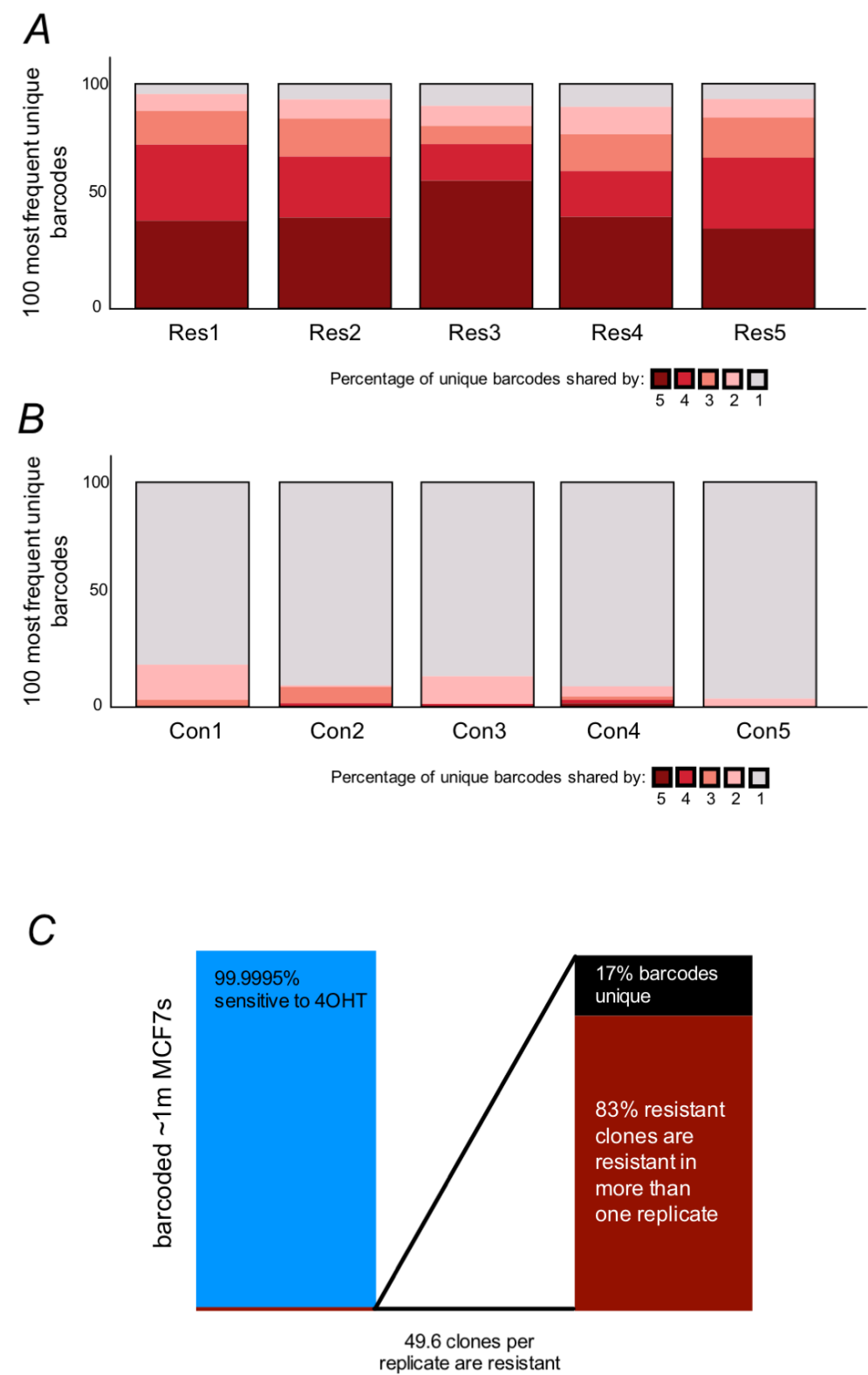
Figure 3.7.2.2 – Shared barcodes in resistant cell line replicates**Figure 3.7.2.2** – Shared barcodes in resistant cell line replicates

Combining Pearson's r scores as Figure 3.7.2.1 but for resistance cell lines. Results show reproducibly high correlations between replicates of resistance lines and therefor likely clonal selection of the same cellular progeny. Mean Pearson's r scores were 0.4261 (res1), 0.3756 (res2), 0.4824 (res3), 0.3876 (res4) & 0.4675 (res5). All individual pairwise comparisons were statistically significant at least $p < 0.001$ by Pearson's r .

Figure 3.7.3 – Shared barcodes across cell lines

A) Expanding our analysis to top100 most abundant barcodes in cell line shows that the majority of enriched populations are enriched across multiple replicates in MCF7-CT res lines. Specifically, an average of 42% of the most frequent barcodes in one cell line were also enriched in each of the other four. Only 8% of the most frequent barcodes were uniquely enriched in that cell line only. **B)** In control lines, an average of 86% of barcodes in the top 100 enriched in one cell line unique to that cell line. <1% of barcodes were enriched across more than two control lines. Hence, the majority of enriched barcodes are unique to each individual replicate. **C)** Graphical depiction of our analysis. In each cell line, 49.6 barcodes were seen (RBA in top100) in all five technical replicates, of these, 83% were found enriched (i.e. RBA in top100) in at least one other cell line. This differs from (A) in that selected barcodes must be present in the top100 across all replicates, not simply have a cumulative RBA in the top100 for that line. **Figure Overleaf.**

Figure 3.7.3 – Shared barcodes across cell lines



3.8 Dynamics of Resistance Over Time

Experiments in this chapter, have shown the reproducible selection of clonal populations in the emergence of 4-OHT drug resistance in an MCF7 cell line model. In this section, we first investigate genomic selection in our resistance models, before reviewing the dynamics of barcode selection.

The development of tamoxifen resistance has been linked to reprogrammed FOXA1 binding patterns (238), the emergence of an EMT profile (319,366), upregulation of *ERBB2* (272) and the development of mutations in ER α itself (345,367). Hence, at least in some cases, tamoxifen resistance could be linked to genomic clonal selection. Several computational approaches have been developed to cluster genetically similar cells into putative genomic clonal populations (for example CloneHD (368) or PyClone (195)). To investigate genomic clonal selection in our model of tamoxifen resistance, we performed exome sequencing in treatment naïve MCF7s, and the same cells after 2 weeks, 3 months and 6 months 4-OHT treatment, in the same resistance protocol as shown in *Figure 3.3.1*. *Figure 3.8.1 A* shows unfiltered VAF plots for two biological replicates of each time point during the derivation of resistant sub-populations. *Figure 3.8.1 B* presents these data as resistant (*i.e.* 6 months treatment) versus sensitive (*i.e.* MCF7 parental) with bar heights representing the range across biological replicates. In either case, the data do not appear to support any evidence of selection for specific resistance conferring SNVs. Indeed, MCF7s appear to be relatively genomically stable during long term culture. Next, we applied PyClone, a Bayesian clustering method for grouping sets of somatic mutations into putative clonal clusters. PyClone estimates cellular prevalence of clones and accounts for allelic imbalances introduced by segmental copy number changes (199). This analysis does not appear to support evidence for genomic selection during the

development of tamoxifen resistance in our models³⁵ (*Figure 3.8.1 C*). Interestingly, these data do support that a long-established cell line is capable of maintaining several stable genomic clusters. From these results, we can putatively assume that reproducibly selected barcodes predominantly mark isogenic but functionally distinct cellular populations, with a predisposition toward becoming resistant.

Lentiviral lineage tracing has previously been applied to CRC PDX models and used to study the development of resistance to chemotherapy (63). In this study, human CRC cells were tagged by lentiviral integration and passaged through multiple generations of PDX model, before being treated with a chemotherapeutic agent. By following lentiviral integration sites, the authors were able to identify individual cellular clones and describe their growth patterns. Of the five growth patterns described, ‘*Type IV*’ clones were defined as being below the level of detection in some passages but reappearing in the tumour later. The authors describe these *Type IV* clones as being quiescent in nature and found them particularly likely to repopulate the tumour after chemotherapeutic treatment (196).

Although such resistant phenotypes are arguably more likely to be observed when studying cytotoxic, cell-cycle stage-dependent agents, we sought to investigate the clonal dynamics of tamoxifen resistance in our barcode model. Hence, we removed excess cells from routine passage at the 3-month timepoint during the development of resistance (T-3months) and analysed the barcode abundance in comparison to that of the final 6-month timepoint (T-6months). *Figure 3.8.2.1* shows average RBA in MCF7-CT control cell lines and *Figure 3.8.2.2* resistant cell lines. In each case, RBA was averaged across technical replicates, and the 100 most abundant barcodes in the final time point (T-6months) were analysed. Interestingly, in each resistant line, the ~25% most enriched barcodes at 6 months, were not in the ~25% most enriched barcodes after 3 months

³⁵ It is worth noting that exome sequencing, by definition, will not identify SNVs in the non-coding genome (for example, mutations in the promotor or enhancer regions of genes). We cannot rule out selection of a genomic clone defined by such somatic variations.

(*Figure 3.8.2.2*). This was not the case in sensitive cell lines (*Figure 3.8.2.1*) where abundance of specific barcodes at 3 and 6 months was relatively stable. It appears from these results that the eventual dominant resistant clones are not the same clones with an initial proliferative advantage when cultured with 4-OHT. This could suggest the development of a delayed-onset resistance phenotype or simply reflect the stochastic, time-point-dependent, dominance of certain clones in the resistance model.

Further sampling timepoints are required to further delineate the contribution of quiescent or *Type IV* clones in the development of tamoxifen resistance in our models. However, as a final experiment, we sought to identify barcoded populations that were relatively quiescent in control replicates in T0-months to T3-months and uncover the contribution of these ‘quiescent-like’ cells to the final dominant pool of resistant cells in our MCF7-CT res cell lines. *Figure 3.8.3* shows the results of this analysis; we defined ‘quiescent’ clones as those which were not part of the top200 most frequently seen barcodes in any of the control lines. Of these, we ranked based on prevalence in the final (6 month) resistant pool. 100 of such barcodes are shown in *Figure 3.8.3 A*, although our analysis selects for these phenotypes, it is clear that relatively quiescent cells are contributing to the development of resistance to tamoxifen. Moreover, if we consider the most frequent 100 barcodes seen at T-3months and T-6months in our treatment arm and analyse their corresponding abundance in control lines (*Figure 3.8.3 B*) we find significantly higher prevalence of T3-month barcodes in the control lines than at T6-months ($p < 0.0001$). This suggests that relatively slower proliferating cells are responsible for the eventual dominant resistant clones. Those with early resistance to tamoxifen, are at least in part, fuelled by faster growth in even control conditions.

Figure 3.8.1 – VAFs and PyClone in development of resistance to 4-OHT

*Analysis of genomic selection in our MCF7 resistance model. **A)** VAFs of MCF7 SNVs during the resistance process. Lines denote individual SNVs sampled at different time points. Each time point consists of two independent replicates of exome sequencing. **B)** Data as in (A) but focussing on 6-month culture versus parental (untreated). Differences between red and blue bars (range over 2 replicates) suggest SNV enrichment. None is seen. **C)** PyClone statistical inference of clonal structures (SNV number contributing to each cluster in legend), again no noticeable clonal selection is seen. **Figure Overleaf.***

Figure 3.8.1 – VAFs and PyClone in development of resistance to 4-OHT

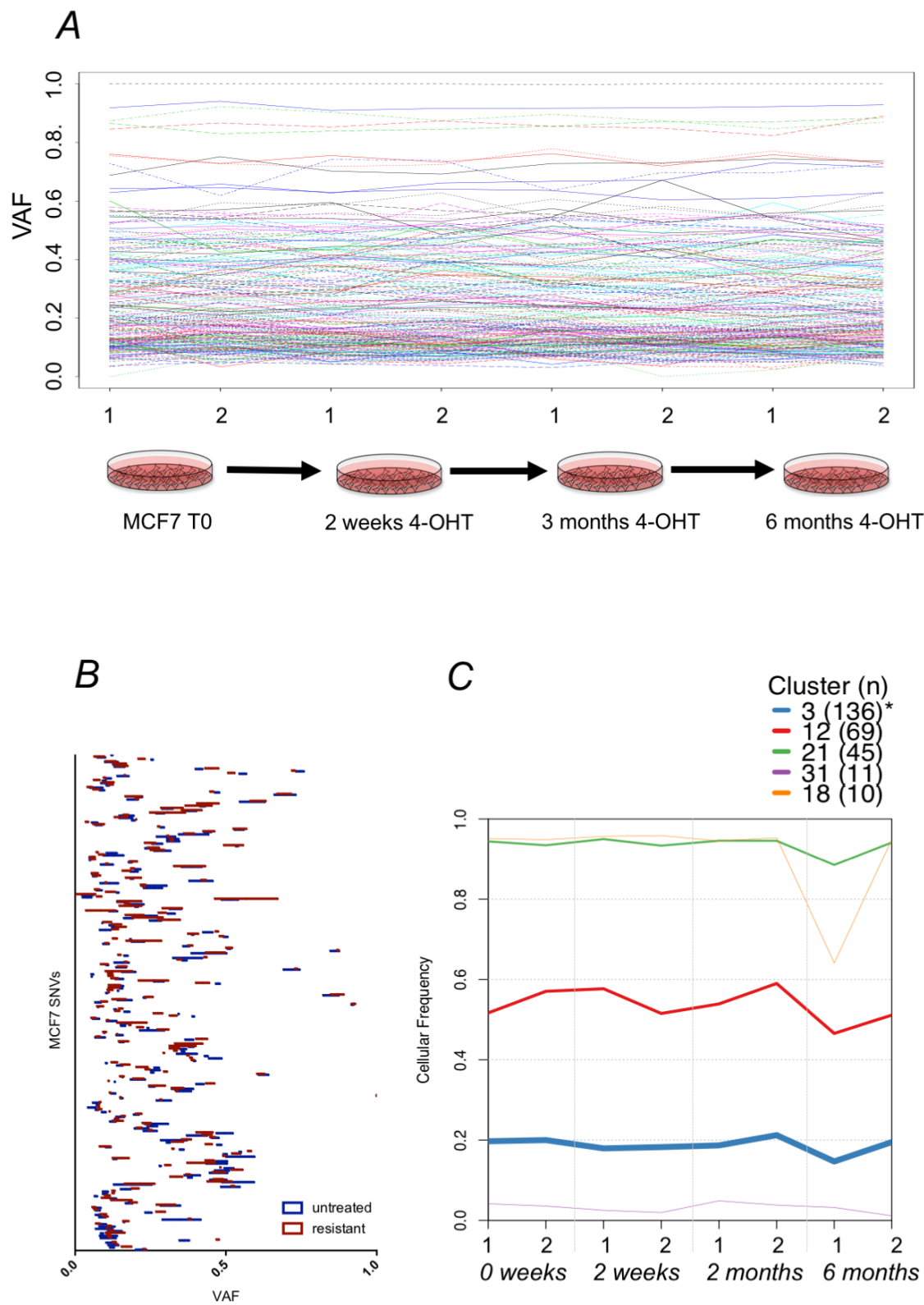
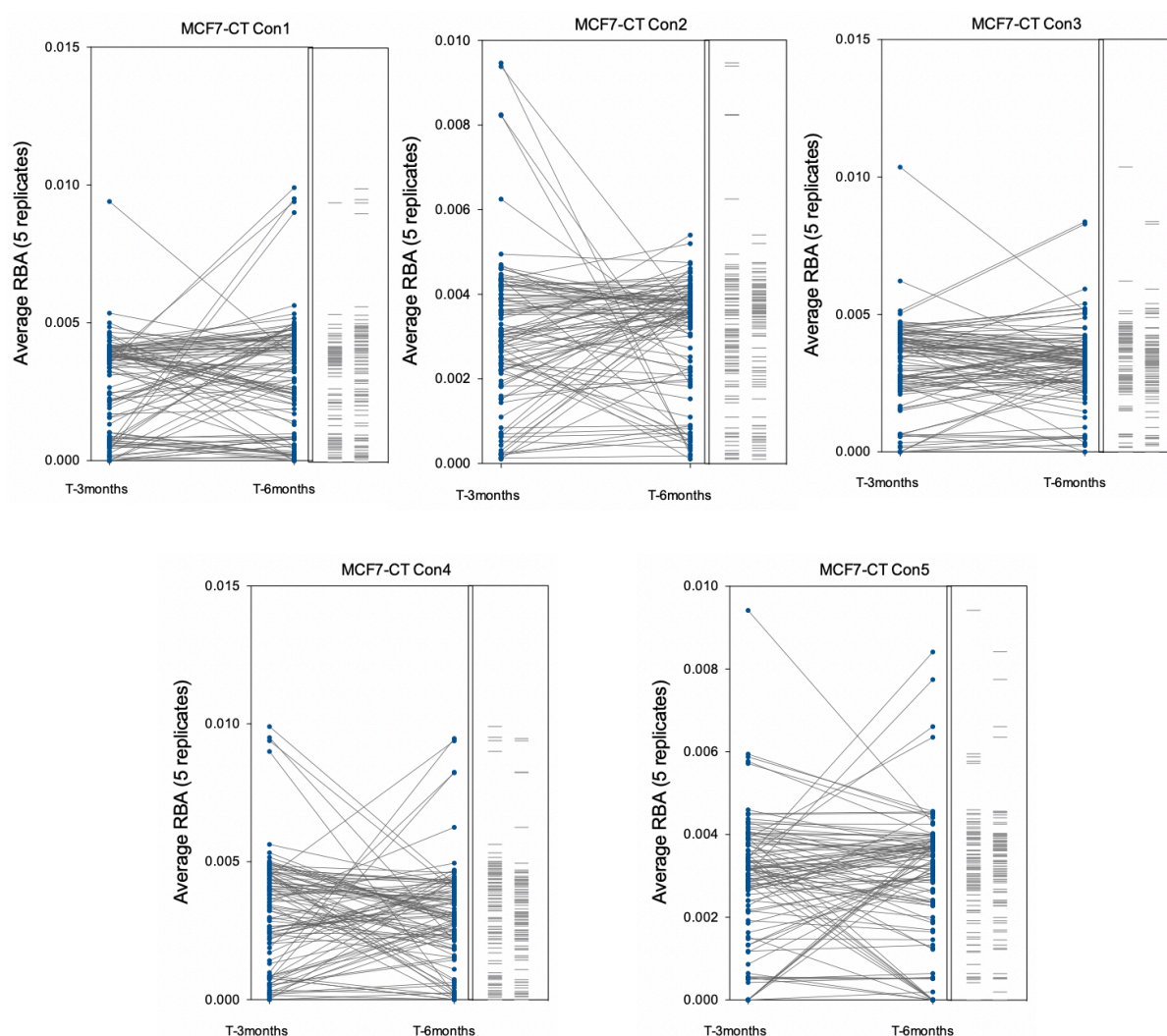
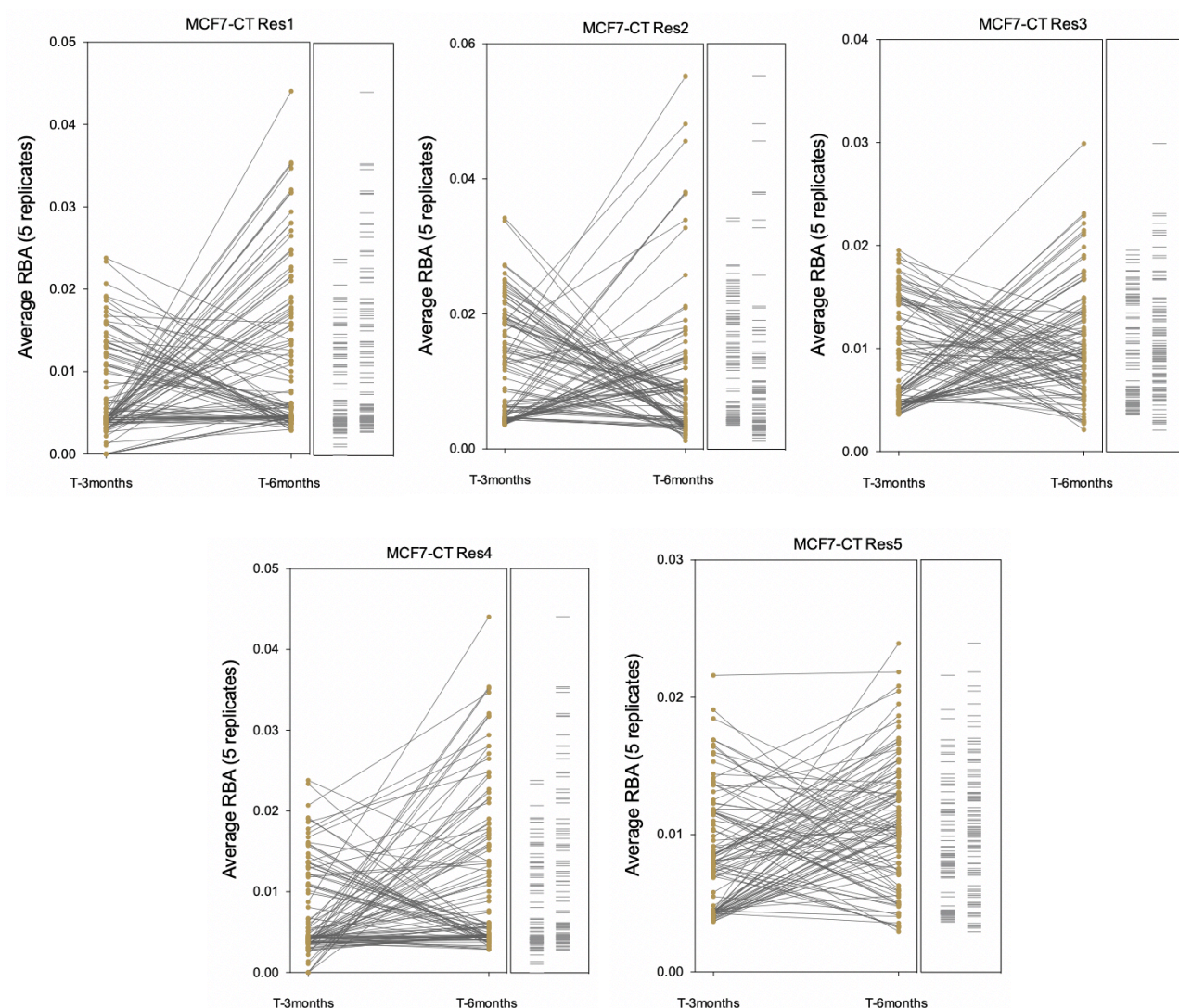


Figure 3.8.2.1– Dynamics of barcode abundance in control lines**Figure 3.8.2.1**– Dynamics of barcode abundance in control lines

Samples were taken during the development of tamoxifen resistance at T3-months culture and again at T6-months and barcode abundance analysed by NGS. RBA was averaged across technical replicates, and the 100 most abundant barcodes in the final time point (T-6months) were analysed. Average fold change for each cell line was: con1 = 1.78, con2 = 2.05, con3 = 1.03, con4 = 1.88 & con5 = 0.99.

Figure 3.8.2.2 – Dynamics of barcode abundance in resistant lines**Figure 3.8.2.2** – Dynamics of barcode abundance in resistant lines

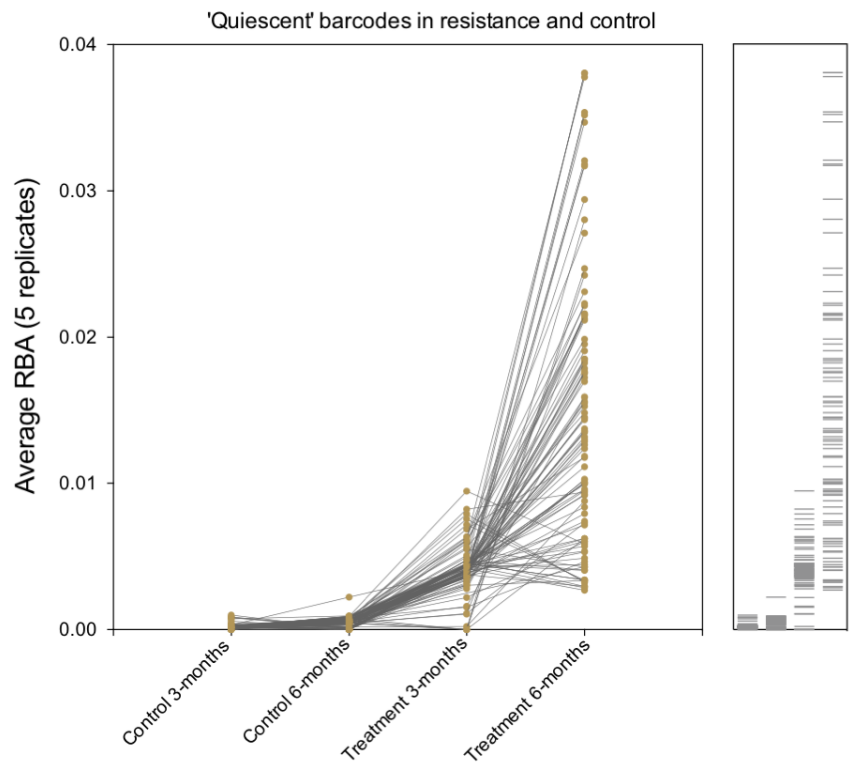
As Figure 3.8.21 but showing resistance replicates. Results appear to show that resistance barcodes have a bimodal frequency, with the dominant (most abundant) at T3-months typically being less dominant at T6-months and vice versa (by visual analysis). Average fold change for each cell line is similar to controls: res1 = 2.52, res2 = 1.58, res3 = 1.63, res4 = 2.29 & res5 = 1.61.

Figure 3.8.3 – Quiescence in resistant cell lines

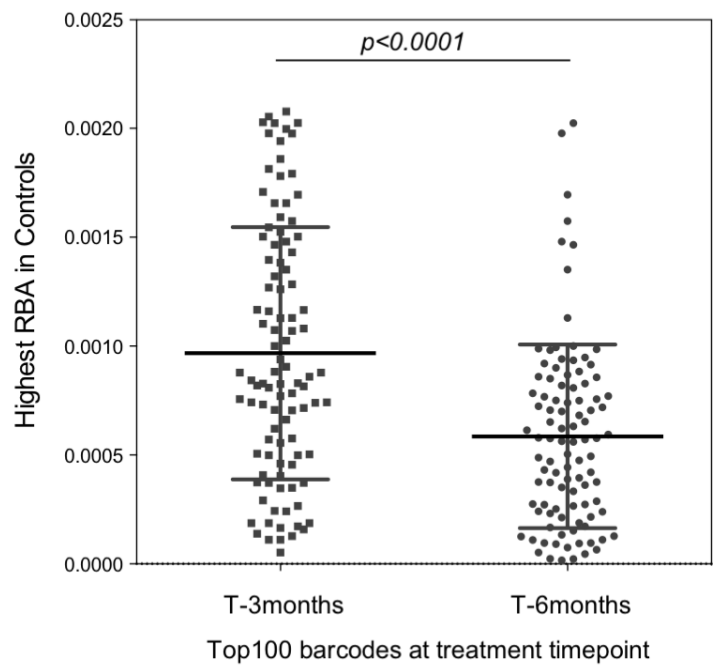
*This analysis seeks to investigate the contribution of quiescent or slowly dividing cells to the development of resistance. **A)** Barcodes which were not part of the top200 most frequent in any control lines were selected as putatively quiescent. They were then ranked based on their presence in the final resistant cell line mixed pool and the top100 plotted over time. Clearly some slow growing cells under normal conditions are contributing to the development of resistance. **B)** Highest RBA in controls for the top100 barcodes enriched across resistance lines at 3 and 6 months. Significantly more of these slow growing cells contribute to the 6-month pool than 3-month pool (Mann-Whitney $p < 0.0001$). **Figure Overleaf.***

Figure 3.8.3 – Quiescence in resistant cell lines

A



B



3.9 General Discussion

In this chapter, we apply lentiviral lineage tracing technologies to a well characterised model of breast cancer drug resistance, that of the outgrowth of MCF7 cells exposed to long-term oestrogen free culture in the presence of 4-OHT. Firstly, we designed and optimised a reproducible method of producing tamoxifen resistant derivatives of the MCF7 line. Previous studies have derived tamoxifen resistant cell lines using a variety of protocols. For example, *Leung et al.*, established five separate lines, either the presence of increasing concentrations of tamoxifen (TamR3, TamR6 and TamR7) or in the absence of oestrogen (TamC3 and TamC6). TamR7 cells were additionally cultured in the presence of oestrogen (345). Large differences in cell lines were uncovered in DNA content (ploidy), mode cell volume and cell cycle time. Cell lines could be further classified based on drug sensitivity. Rapamycin alone or in combination with tamoxifen increased the level of phospho-Akt in parental MCF7 cells and TamR7, TamC6 and TamR6. Increases in phospho-Akt in some cells are due to an inhibitory feedback mechanism between the mTOR effector p70 S6K and the insulin receptor substrate- PI3K upstream of Akt (369). Although difficult to profile *in vitro*, culture conditions used in this Chapter (lack of exogenous oestrogens and increasing concentration of 4-OHT) would likely result in an oestrogen independence phenotype, characterised by lack of sensitivity to both 4-OHT and AIs. Resistant cell lines derived through our protocol were characterised by a ~10-15x increased IC₅₀ for 4-OHT, ER α expression but a lack of transcriptional response to exogenous ER α ligands (slight downregulation of ESR1 & PGR expression in each case), decreased baseline proliferation and lack of proliferative response to either E2 or 4-OHT (*Figures 3.2.1 & 3.2.2, 3.3.1*).

Interestingly, there is some heterogeneity in oestrogen signalling amongst resistant cell lines described in the literature. In the example by *Leung et al.*, lines cultured in the absence of oestrogen either tended to maintain high *ESR1* expression (and ER α responsiveness) accompanied by an increase in HER2 protein levels, or to slightly reduce

ER α expression and compensate by increased phosphorylation of EGFR (345). Resistant lines cultured in the presence of oestrogen (with tamoxifen), in contrast, tended to decrease ER α expression. ER α shows greater binding affinity for 4-OHT than E2 in most cases, so it is puzzling that the presence of both ligands would significantly alter the resistance process (370). Interestingly, in each case of our resistant lines, we observed an increase in the steepness of the Hill Slope of dose-response curves. This is typically a measure of cooperative binding, so it is possible that ER α affinity is changed during the resistance process (371). To our knowledge, this is the first time this feature of tamoxifen resistance has been reported. Competitive ligand binding studies would be needed to further elucidate and clarify the observation (372).

Heterogeneity in ER α + breast cancer, reflected in histology (373), genetic architecture (7,73) and transcriptional regulation (374), is common place and is thought to ultimately impact on the long-term response to endocrine therapy (375). Despite this, recent studies in ER α + (luminal) breast cancers have found that driver coding mutations do not significantly change between primary and metastatic sites, with the notable exception of ESR1 mutations (157), suggesting that alternative non-genetic mechanisms may contribute to progression and resistance to endocrine therapy (157,376). Indeed, epigenetic modifications have been shown to modulate ER α binding to enhancers, for example by interacting with ER α -associated pioneer factors (236,377). In addition, a recent study from *Hinohara et al.*, found that genetic deletion or inhibition of the KDM5 histone H3 lysine 4 demethylase family increased sensitivity to anti-oestrogens by modulating ER α signalling and decreasing cellular transcriptomic heterogeneity (483).

Increasing evidence suggests that epigenetic information can actively transfer gene transcription states across cell division (378–381), delineating a plausible mechanism by which epigenetic clonal evolution could give rise to treatment resistance in cancer. A recent study by *Patten et al.*, revealed several critical principles underlying phenotypic heterogeneity in breast cancer progression (376). By comparing samples from endocrine

therapy resistant metastatic patients with treatment-naïve primary samples, the authors uncovered a set of enhancers marking phenotypic clones that significantly expanded during disease progression. The authors go on to present evidence that progressive activation of FOXA1 and its network in metastatic samples is a consequence of expansion of a phenotypic clone. This cellular clone was marked by an active FOXA1 enhancer (376); FOXA1 is a well characterised modulator of ER α - chromatin interactions as well as a critical determinant of ER α -transcriptional regulation and endocrine response in breast cancer (238). These studies present a viable mechanism by which evolution of phenotypic, or epigenomically regulated, clonal populations could contribute to the rapid development of endocrine resistance in the >40% of ER α breast cancer patients who ultimately relapse during or after completion of adjuvant endocrine therapy (382)

Further evidence for the functional role of FOXA1 in resistance to endocrine therapy comes from a study by *Ross-Inness et al.*, (248). In this study, the authors mapped genome-wide ER α -binding events in primary breast cancer by chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq). They found that in patients with poor clinical outcomes, or in those with distant ER α -positive metastases, ER α was still recruited to the chromatin but that new ER α -binding regions, unique to poor prognosis, were acquired. Interestingly, the authors found that these new binding regions were driven by the ER α pioneer factor, FOXA1, and that FOXA1 binding could be altered by 90 minutes treatment with a mitogenic cocktail (containing epidermal growth factor (EGF), insulin-like growth factor-1 (IGF-1), interleukin-6 (IL-6) and tumour necrosis factor alpha (TNF- α)). The authors conclude that the differential ER α -binding programme observed in tumours from poor prognosis patients was not due to the selection of a rare subpopulation of cells, rather a FOXA1-mediated reprogramming of ER α binding on a rapid timescale (248).

More broadly, genomic intratumour heterogeneity is thought to drive a majority of resistance processes. Within the concept of population genetics, cancer cells are

genetically unstable and this instability drives heterogeneity in even established tumours (383). Indeed, the wider the diversity of minor clonal populations in a tumour, the more likely it is that resistance will arise. Such an association between tumour heterogeneity and drug resistance has been noted in ovarian (136), and oesophageal (137) cancers. Additionally, basal-like TNBCs have previously been linked with shorter disease free survival compared to non-basal-like TNBCs and tend to be associated with higher clonal diversity (78). Thus, cancer has been described as a ‘moving target’ as the population of cells constantly shifts during therapy (384). A better understanding of the evolutionary process that underlies drug resistance is needed for avoiding or postponing the emergence of resistance.

NGS has been heavily relied on in studies to elucidate the clonal origins of drug resistance. However, it is limited to a sensitivity of 0.1% allelic fraction (385). Given that the detectable tumour burden is estimated to be approximately 10×10^9 tumour cells at the time of diagnosis (134), this level of resolution is clearly insufficient to comprehensively assess pre-existing cancer subpopulations. We reasoned that cellular barcoding, which has been used to trace lineage during hematopoietic stem cell differentiation (201–203), could be used to address this question and overcome the limited sensitivity of current NGS approaches. To trace the origins of resistant populations in our model of tamoxifen resistance, we opted to use the high complexity barcode library, ClonTracer, (see *Figure 3.4.1*). These molecular barcodes enable the labelling and tracking of >10 million individual cells with a unique 30-nucleotide long semi-random DNA sequence tag (135). The ClonTracer library was designed to have a balanced GC content to ensure uniform PCR-amplification efficiency and has a theoretical complexity of 73 million unique barcodes (135). Having established a protocol for the derivation of resistant cell lines, we sought to integrate complex molecular barcodes in the ClonTracer platform via lentiviral integration in unchallenged MCF7 cells.

Figure 3.5.1 shows the experimental design of our resistance screen. Plating was such that resistant and control replicates of the same number (*e.g.* MCF7-CT res1 and con1) were derived from the same originating Eppendorf, allowing direct comparisons to be drawn. To accommodate the need for 100 million cells per replicate, biological replicates (5 in each arm) consisted themselves of 5 x 15cm dishes (seeding density of 4×10^6). These technical replicates were maintained and expanded separately. Thus, 5x resistant and control pairs (consisting of 5x technical replicates) were established and propagated. Resistance phenotypes were as expected from our optimisation steps (*Figures 3.5.3.1, 3.5.3.1, 3.5.4 & 3.5.5*). Interestingly, resistant lines were found to be relatively insensitive to exogenous E2 (in terms of ESR1 and PGR transcription), in keeping with the assumption that long term culture in the absence of E2 (alongside exogenous 4-OHT) could lead to oestrogen independence. Further studies should confirm whether our MCF7-CT res lines are in fact doubly resistant to both AIs and SERMs, such as tamoxifen.

The abundance of unique barcodes in each sample were computed, as detailed in *Figure 3.5.2*, and complexity of the library ensured pre and post infection (*Figures 3.4.2 & 3.4.4*). We reasoned that as each originating cell would be uniquely marked by the highly complex barcode library, fitness of cellular clones would be indicated by barcode abundance after drug treatment. Should the same barcodes be enriched in multiple replicate plates, this would indicate the presence of pre-existing resistance populations.

Resistance cell lines universally displayed enrichment for specific barcodes (*Figure 3.6.1 & 3.6.2*), which was not seen in control lines. Interestingly, 14 barcodes were enriched in control lines, though not to the same extent as the resistance lines (*Figure 3.6.3*). Retroviral based therapies have long been known to induce oncogenic transformation in a proportion of patients. It is thought that oncogene activation is caused by viral integration in specific genomic regions (364). Lentiviral based vectors are generally considered safer, they are known to integrate preferentially in megabase-wide

chromosomal regions, and not in activating regions associated with cell growth (365). However, we sought to eliminate any potential cases of oncogenic viral integration in our cell lines and exclude these from our analysis, and so removed these 14 barcodes from both sensitive and resistance cell line analyses (*Figure 3.6.3*).

Importantly, enriched barcodes show a high concordance between both technical replicates and across resistant cell lines (*Figure 3.7.1, 3.7.2.1 & 3.7.2.2*). It seems likely that this is evidence of increased fitness amongst the enriched barcoded clones, rather than stochastic population dynamics. On average, 49.6 barcodes were enriched in in all five technical replicates in resistant lines, of these, 83% were found in at least one other cell line (*Figure 3.7.3*). This enrichment is not accompanied by selection for specific SNVs or genomic clonal populations as previously described (*Figure 3.8.1*) (317), although non-coding SNVs would not have been detected by our analysis. Thus, it appears that, at least in our model, prolonged exposure to tamoxifen results in the expansion of isogenic clonal populations predisposed to tamoxifen resistance. In the majority of reported cases, clonal selection of resistance populations in cancer is associated with the outgrowth of a genetically distinct subclone. For example, the development of resistance to AIs in ER α + breast cancer has been linked to activating mutations in the ER α gene itself (*ESR1*) (386,387). Subclonal activating mutations are thought to be present in a third of patients with prior AI exposure (388–391), and detection of specific *ESR1* mutations in circulating tumour DNA (ctDNA) has recently been shown as an effective early biomarker of AI resistance (392). Our results are relatively novel in showing cellular clonal selection without accompanying genomic changes in a well-studied resistance process.

Recently, *Hinohara et al.*, explored the role of KDM5 histone H3 lysine 4 demethylase family in the development of endocrine resistance (483). The authors found that deletion of KDM5A/B or inhibition of KDM5 activity increased overall sensitivity to anti-oestrogens through modulation of ER α signalling. Further, the authors uncovered a

prognostic KDM5 expression signature in ER α + human breast cancers and demonstrate a link between KDM5B expression and transcriptomic heterogeneity. Crucially, molecular barcoding of MCF7s using the ClonTracer system (135) revealed that development of both tamoxifen and fulvestrant resistance was due to the selection of a rare, pre-existing, population of cells. Though only using four replicates for each condition, the authors present evidence that a great deal of overlap existed between each population.

Finally, in this *Chapter*, we sought to profile the dynamics of the development of tamoxifen resistance. Quiescent cellular populations have been described to drive resistance to cytotoxic therapies in at least one model (83). Because of limited sampling points in our experiments, we opted to define quiescent clones based on barcode abundance in control lines, before looking to the contribution of these populations in our resistance models. There was a clear tendency toward early dominant populations (those with highest abundance at 3 months) to be replaced by a slower growing population after full resistance was developed (*Figure 3.8.2.2*). Interestingly, not only did quiescent cells in the control lines appear to have a large influence on resistance in the treatment arm (*Figure 3.8.3 A*), there appears to be an early resistance phenotype driven by relatively proliferative cells (*Figure 3.8.3 B*) but eventually replaced by these slow cycling cells.

Overall, in this *Chapter* we have defined and optimised a method for developing resistance to tamoxifen, and potentially complete oestrogen independence, in the MCF7 cell line. We have shown that resistance is typically the result of the clonal expansion of isogenic cellular populations with a predisposition to resistance. These populations may tend toward quiescence in normal culture conditions, though more work is required to fully investigate this point.

Chapter 4

The Role of BC-TICs in Tamoxifen Treatment & Resistance

Aims of the chapter

1. Profile the tamoxifen resistance models for BC-TIC signature.
2. Characterise the BC-TIC response to acute tamoxifen treatment.
3. Investigate the role OCT4 in tamoxifen resistance and the BC-TIC phenotype.

4.1 Introduction

Numerous routes to, and causes of, tamoxifen resistance have been proposed in the literature. These generally fall into two broad categories; activation of alternative oncogenic signalling pathways, such as PI3K/EGFR/HER2 (258,266,344,345), typically resulting in Akt activation, or reprogramming of ER α recruitment to the chromatin, involving PAX2/GATA3/FOXA1 (236,238,272). These mechanisms are not mutually exclusive, with at least one major study relating PAX2-related ER α reprogramming to enhanced ERBB2 expression (272). A third mechanism, reported by numerous studies but remaining controversial, is the activation of pluripotency factors and induction of EMT (209,242,277,282,393–395).

ER α and its related pioneer factors play a central role in stem cell differentiation and maintenance in the normal human mammary gland (236,237). Indeed, it is these pioneer factors which determine the specific cellular context dependent role of oestrogen. For example, FOXA1 is highly expressed in luminal cells and has an essential role in formation of the terminal end bud (TEB). Loss of *GATA3* similarly results in severe defects in mammary development due to failure of TEB formation. Furthermore, *GATA3* has been shown to participate in luminal epithelial differentiation and lobuloalveolar development (234,237,239). Tamoxifen has complex actions on the ER α , similarly linked to the various coregulatory proteins controlling ER α tissue specificity and transcriptional programs. For example, in the breast, it acts primarily as an antagonist or biased agonist, but in other tissues such as the uterus, cardiovascular system and bone, it acts as a full ER α agonist (267).

At least one study has shown directly that oestrogen reduces the overall proportion of stem cells in the normal gland (276). Embryonic transcription factors *NANOG*, *POU5F1*, and *SOX2* expression decreases upon differentiation and with oestrogen treatment, and similarly overexpression of any of the three reduces *ESR1*

expression, and increases the number of stem cells in the normal mammary gland (276). Given the role of ER α in the developing mammary gland, and the direct regulation of embryonic transcription factors by oestrogen, it is probably unsurprisingly that resistance to ER α inhibitors has been linked to dedifferentiation. All three canonical factors required for iPSCs have been variably linked to tamoxifen treatment or resistance. For example, tamoxifen resistant cells are enriched for mammosphere forming cells and show higher expression of *SOX2*. Silencing of *SOX2* by siRNA reduces the stem/progenitor cell population and restores sensitivity to tamoxifen (277). Secondly, OCT4 expression was shown to be independently associated with poor prognosis in 319 cases of invasive breast cancer (281), and has recently been described as a novel ER α -associated transcription factor involved in ER α recruitment to tamoxifen, but not oestrogen, associated transcriptional sites (282). Conversely, KLF4 is generally associated with a favourable outcome in breast cancer (280,284). *Jia et al.*, found that higher expression of *KLF4* correlated with increased tamoxifen sensitivity in patients, and was positively correlated with ER α activity (284).

More broadly, the development of tamoxifen resistance has been linked to the BC-TIC phenotype and the processes associated with EMT (366). BC-TICs are thought to be slow cycling and highly expressing ABC multidrug resistant transporter proteins; both enriched following cytotoxic chemotherapy and endowed with inbuilt resistance (181,183). This phenotype has been associated with upregulation of *ALDH1A1* (166,396), downregulation of *miR-375* (397), downregulation of *FOXA1* with induction of *IL6* (398) and countless other mechanisms (104,149,187,188,399). Numerous studies have also linked breast cancer drug resistance to EMT (104,393), and EMT to the acquisition of a BC-TIC-like phenotype (394,400), though our own studies have found that these qualities are largely separable (159).

Recently, the DNA methylome of endocrine sensitivity was characterised by *Stone et al.*, (285). The authors found that in tamoxifen resistance, DNA hypermethylation occurred at oestrogen-responsive enhancers. This was associated with reduced ER α -chromatin binding and consequently decreased gene expression of key oestrogen regulators. In another study by this group, promotor demethylation with 5-Aza, coupled with E2, restored ER α -regulated gene expression in TamR lines and induced a significant anti-proliferative effect (97). Moreover, *Patten et al.*, have recently mapped the epigenome of 47 primary and metastatic ER α + breast cancers by ChIP-seq for the active enhancer marker, H3K27ac (376). H3K27ac is a modified Histone H3, associated with higher activation of transcription, and so allowed the authors to build a comprehensive picture of clinically relevant active regulatory regions in clinical breast cancer samples. The authors identified YY1 as a critical determinant of ER α transcriptional activity and a marker of phenotypic clonal populations involved in the development of resistance to endocrine therapies (376). Interestingly, an earlier study by the same authors, linked an invasive, EMT, like phenotype to the development of acquired resistance to AI and to AI-induced epigenomic reprogramming, highlighting the potentially interdependency of an EMT-like phenotype and resistance to endocrine therapies (401).

Cumulatively, these results highlight a novel role for ER α response element methylation, and more broadly phenotypic heterogeneity in the epigenome, in the induction of resistance to endocrine therapies. Interestingly, embryonic transcription factors OCT4, SOX2 and KLF4 each have roles in regulating the epigenome of stem and progenitor cells to control differentiation (286,287). Additionally, the oncogenic histone methyltransferase and key embryonic regulator, EZH2, has been shown to contribute to tamoxifen resistance by silencing the expression of ER α cofactor *GREB1* (288). Together, these results highlight a potential mechanism by which embryonic transcription factors contribute to

regulation of ER α transcriptional activity and hence influence the development of tamoxifen resistance.

Given our previous findings that the development of tamoxifen resistance was not due to selection of a specific genomic cluster defined by coding mutations, but polyclonal in nature (see: *Chapter 3*), we sought to profile further the potential roles of EMT and BC-TIC-like phenotypes in the resistance process. Given the role of ER α in controlling differentiation in the normal human mammary gland, resistance induced pluripotency seems a plausible avenue in explaining the multitude of resistance mechanisms described in anti-oestrogen therapy. Specifically, in this chapter we profile the expression and function of pluripotency related transcription factors in the development of resistance and the maintenance of a BC-TIC phenotype. We use a BC-TIC surrogate assay of mammosphere formation to investigate the contribution of 4-OHT to mammosphere initiating cells (M-ICs). Our analyses uncover a tamoxifen-inducible BC-TIC/M-IC phenotype in cell line and PDX models and we present evidence for an OCT4-related transcriptional program central to both resistance and tumour initiation.

4.2 M-ICs are enriched by 4-OHT treatment

Resistance to tamoxifen has been linked to expression of several master transcription factors typically associated with guiding embryonic stem cell fate and differentiation. We sought to profile expression of these transcription factors alongside putative BC-TIC markers in a panel of ER α + tamoxifen sensitive (MCF7 parental, T47D, MCF7-CT con1-5) and resistance (TamR, MCF7-CT res1-5) cell lines. *Figure 4.2.1* shows the results of a qPCR panel for embryonic transcription factors OCT4 (*POU5F1*), *SOX2*, *NANOG* & *KLF4* alongside markers of BC-TICs *CD44* and *PROCR* (167,402), results are presented as fold-change relative to MCF7 parental, statistical significance (by one-way ANOVA with Bonferroni; * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$) is also presented versus MCF7 parental. Several markers appear to be enriched in resistant models. This is clearest in *POU5F1*, *SOX2* & *PROCR* where resistant and sensitive groups have mean relative expression of *POU5F1*: 3.42 & 1.13, *SOX2*: 3.53 & 1.21 & *PROCR*: 2.79 & 0.93 respectively. In *Figure 4.2.2*, mean expression in sensitive and resistant models are combined. Significant differences are seen between sensitive and resistant groups (by one-way ANOVA with Bonferroni correction) for *POU5F1* ($p=0.0023$), *SOX2* ($p=0.0057$), *KLF4* ($p=0.0015$) & *PROCR* ($p=0.00027$) expression. There is also a slight enrichment in *CD44* expression in the resistant lines, though this is not statistically significant. These results appear to be in keeping with the literature: *KLF4* has been associated with good prognosis and is preferentially enriched in sensitive models (284); *SOX2* & OCT4 have been directly linked to tamoxifen resistance (277,282), and are enriched in our resistant models; surface markers *CD44* & *PROCR* are thought to mark BC-TICs and/or mammary stem cells (164,167,403) and again are enriched in our resistant models.

We next sought to investigate the functional manifestations of enriched BC-TIC markers in our resistant models. Sphere formation assays were first developed to

enable the growth and purification of putative stem cells from single clones in non-adherent, serum free conditions, and have since been applied to study multiple stem cell and phenotypes in normal and cancerous tissues (404,405). In breast cancer research, the mammosphere assay has been heavily relied on as an *in vitro* surrogate for tumour initiation potential. Hence, M-ICs are thought to be the *in vitro* equivalent of BC-TIC (159). Mammospheres show a range of upregulated stem cell-associated signalling pathways (169) and their formation efficiency correlates with the tumorigenicity of their originating tissues, when ranked by LDAs in mouse xenografts (170,185,406). In order to assay both mammosphere proliferation and mammosphere forming efficiency (MFE), we designed a mammosphere culture and image analysis workflow. Briefly, single cells were plated in low adherent conditions with mammosphere promoting media, with or without the last 24hrs culture being in the presence of drug. After 4 days, first generation spheres would form, these could be passaged and a known number of single cells re-plated to form second generation spheres. The number of M-ICs in the first-generation culture could then be calculated as the ratio of 2nd generation spheres formed to input first generation mammosphere cells, also known as the MFE. Phase contrast images were captured and stitched together in ImageJ before automated counting and size discrimination. Formally, we define MFE as the number of 2nd generation spheres >50µm formed as a proportion of plated 1st generation single cells, whilst the absolute number of 2nd generation spheres gives a measure of M-IC number in the 1st generation culture. We reasoned that our assay variation could also give us a measure of proliferative potential, by quantifying mammosphere size.

Figure 4.2.3 A shows the area of mammospheres formed by TamR and MCF7 cell lines, *Figure 4.2.3 B* shows MFE. TamR cells tend to form significantly more ($p<0.05$; two-tailed t-test), but significantly smaller mammospheres than MCF7s ($p<0.0001$; two-tailed t-test). This suggests reduced proliferation, but enhanced M-IC numbers in the tamoxifen resistant line. In *Figure 4.2.3 C* we see that all five replicates of

barcoded resistant cell lines have significantly increase MFE relative to MCF7 parental and MCF7-CT control lines (statistical significance by one-way ANOVA with Bonferroni correction denoted (* $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$). The output of our ImageJ workflow for scoring mammosphere numbers and areas is shown representatively for MCF7 parental and TamR cell lines in *Figure 4.2.3 D*.

We next asked whether acute treatment with 4-OHT would be enough to induce a CSC-like transcriptional profile and/or enrich cultures for M-ICs. *Figure 4.2.4 A* shows expression of *POU5F1*, *SOX2* and *PROCR* in MCF7s and TamRs cultured for 24hrs in a variety of conditions. In MCF7s, 4-OHT induced expression of all three markers significantly ($p<0.05$), E2 had little effect and a combination of E2 and 4-OHT had a similar affect to 4-OHT alone. As we have seen in *Figure 4.2.1*, models resistant to 4-OHT tend to have a higher baseline expression of all three markers. However, the addition of either ligand, in this experiment, served to significantly reduce TamR expression of CSC markers ($p<0.001$ for *SOX2* & *PROCR*; $p<0.05$ for *POU5F1* (only significant with E2)). Interestingly, increased *PROCR* expression in resistant versus control lines appeared to be relatively inconsistent between experiments (42-fold induction in *Figure 4.2.4 A* and 3-fold in *Figure 4.2.1*. This variability prompted us to exclude *PROCR* from future experiments. Functionally, treating first generation mammospheres with 4-OHT leads to significantly reduced size mammospheres in the 2nd generation ($p<0.001$), but significantly increases the number of M-ICs ~3.5 fold (*Figure 4.2.4 B*; $p<0.01$). The effect of 4-OHT on mammosphere size appears to follow a dose-response relationship: *Figure 4.2.4 B* shows a stepwise reduction in mammosphere size at higher doses of 4-OHT. We found smaller mammospheres overall in the tamoxifen resistant line (TamR), but also an increased baseline MFE. In this experiment, 4-OHT addition did not appear to alter MFE to a great degree (*Figure 4.2.4 B*).

We hypothesised that as *POU5F1* (OCT4) expression was relatively stable in TamRs treated with 4-OHT, it may hold more significance to the BC-TIC phenotype. Mechanistically, there are several oestrogen response element (ERE) binding sites in a validated promotor region of *POU5F1* (*Figure 4.2.5 A*) (407). Further, this site overlaps with both an ER α /FOXA1 binding site previously associated with the development of tamoxifen resistance and poor outcome in ER α + breast cancer patients (248) and with a tamoxifen-ER α specific site identified by *Welboren et al.*, (408). This second study used ChIP-Seq to map genome wide ER α -binding sites and RNA polymerase II (RNAPII) occupancy in MCF7s. The authors found that tamoxifen and fulvestrant partially reduced ER α binding and RNAPII loading, but had different effects on genes downregulated by E2. The authors conclude that a partially antagonist-loaded-ER α (*i.e.* with tamoxifen and E2) can act as an full agonist at some sites (408). Though this conclusion has since been called into question by *Hurtado et al.*, (238), even this study found that tamoxifen induced ER α binding in a subset of sites induced by E2, albeit weaker in binding intensity. We sought to understand the timing of *POU5F1* induction by 4-OHT treatment and so carried out a time-course qPCR experiment (*Figure 4.2.5 B*). MCF7 monolayers were treated in 24-well plates with 100nM 4-OHT and three wells harvested periodically over a total of 64hrs. Each timepoint was assayed for gene expression by qPCR. Alongside *POU5F1* (OCT4), we assayed three genes thought to be strongly associated with E2-ER α only binding events; *PGR1* (PR), *GREB1* and *WISP2* (238). We found strong induction of *POU5F1* expression from 24hrs, with a corresponding decrease in *PGR1*, *GREB1* and *WISP2* expression (*Figure 4.2.5 B*). It is important to note that only ChIP-qPCR (ChIP for ER α bound 4-OHT, qPCR for *POU5F1*) can confirm direct regulation of *POU5F1* expression by tamoxifen/4-OHT bound ER α . However, these results point to a plausible mechanism by which *POU5F1* expression may be directly induced by 4-OHT bound ER α , at least in some cases.

Interestingly, we also find that mammospheres tend to have a higher tolerance for 4-OHT than monolayer grown cells. *Figure 4.2.6 A* shows an IC₅₀ of 0.85µM for MCF7s grown as spheroids, more than three times the 0.26µM of monolayer cells ($p < 0.0001$; Extra sum-of-squares F-test). This is a feature specific to M-ICs, rather than any nuances of drug penetration in 3D structures. We find 24hrs treatment with 0.6µM 4-OHT markedly effects the viability of MCF7s grown in Matrigel as acinar structures (*Figure 4.2.6 B*). These assays are typically used to study branching morphogenesis in normal breast cell lines (*e.g.* MCF10A) or breast stem cells, but can be a useful 3D model to study cancer cell invasion and growth *in vitro* (327). This result strongly suggests that mammosphere culture enriches for specific cell types with an inherent resistance to 4-OHT, which are not enriched for in either monolayer culture or in 3D growth and morphogenesis assays.

It has been suggested that chronic exposure to anti-endocrine therapy may impart specific selective pressures on breast cancer cells that are ultimately able to induce a stable change in cellular phenotype (401). For example, resistance to AI has been linked to the stable upregulation of genes involved in the 27-hydroxyl-cholesterol biosynthesis pathway which consequently promote oestrogen-independent ER α -chromatin binding at putative regulatory regions (401). It is conceivable that similar, drug-induced, transcriptional changes are occurring in our 4-OHT treated MCF7s. Thus, from the results presented, there are at least two plausible mechanisms: either ER α bound 4-OHT is directly increasing the absolute number of M-ICs, perhaps through induction of *POU5F1* expression, or 4-OHT treatment is selectively toxic in the non-M-IC population, meaning that treatment causes *de facto* enrichment for M-ICs. A schematic of two possible scenarios is presented in *Figure 4.2.7*. We reasoned that if the same population of cells were responsible for resistance to 4-OHT and comprised the bulk of M-ICs, *Figure 4.2.7 A*, then we would see the same ClonTracer barcodes enriched in 4-OHT resistance and in mammosphere culture. A similar scenario would occur if a subpopulation of MCF7s

were predisposed to become resistant, *i.e.* if a 4-OHT induced process led to resistance in only a certain fixed treatment naïve phenotype. Alternatively, as in *Figure 4.2.7 B*, 4-OHT could induce the same process across all cell phenotypes (perhaps the induction of *POU5F1*), which enables some degree of transcriptional plasticity and the (relative) stochastic development of resistance by any cell in the population. This transient state of enhanced plasticity may enable more efficient mammosphere formation, or the final resistant cell type may share properties of M-ICs. In either case, in this second scenario, we would be more likely to see poor correlation between barcodes enriched in resistant cell lines and sensitive cell lines cultured as mammospheres.

Hence, we turned to our MCF7 sub-lines barcoded with the ClonTracer lentiviral cell tracking system. Optimisation and derivation are detailed extensively in *Chapter 3*. We prepared amplicons spanning unique ClonTracer sequences from MCF7-CT con (sensitive to 4-OHT) and MCF7 res (resistant to 4-OHT) 2nd generation mammospheres (shown in *Figure 4.2.3 C*). *Figure 4.2.8 A* shows summary data for correlations between barcodes enriched in MCF7-CT res cell lines described in *Chapter 3* and MCF7-CT con cell lines grown as mammospheres, *Figure 4.2.3 A*. *Figure 4.2.8 B* shows the same but for MCF7-CT con cell lines described in *Chapter 3* and MCF7-CT res cell lines grown as mammospheres, *Figure 4.2.3 A*. Interestingly, we see good overlap between cellular populations (barcodes) enriched in M-ICs in each control line and those enriched in 4-OHT resistance from our studies in *Chapter 3*. These data suggest that 4-OHT is selecting for pre-existing cellular populations which are enriched for M-ICs. Interestingly, it is possible that such a population may be ‘primed’ for drug-induced transcriptional changes, rather than truly pre-existing prior to treatment (401). However, it is clear that it is M-ICs that ultimately dominate in 4-OHT resistant cell lines.

Figure 4.2.1 – qPCR for stem cell markers

*Stem cell transcription factors NANOG, POU5F1, SOX2 and KLF4 are profiled across tamoxifen resistant and sensitive models by qPCR. Putative surface markers PROCR and CD44 are also profiled. Results are normalized by comparative C_T method ($2^{-\Delta\Delta C_T}$) (333) to the geometric mean of B2M and GAPDH, and MCF7 expression (see Methods, section 2.4.1). **Figure Overleaf.***

Figure 4.2.1 – qPCR for stem cell markers

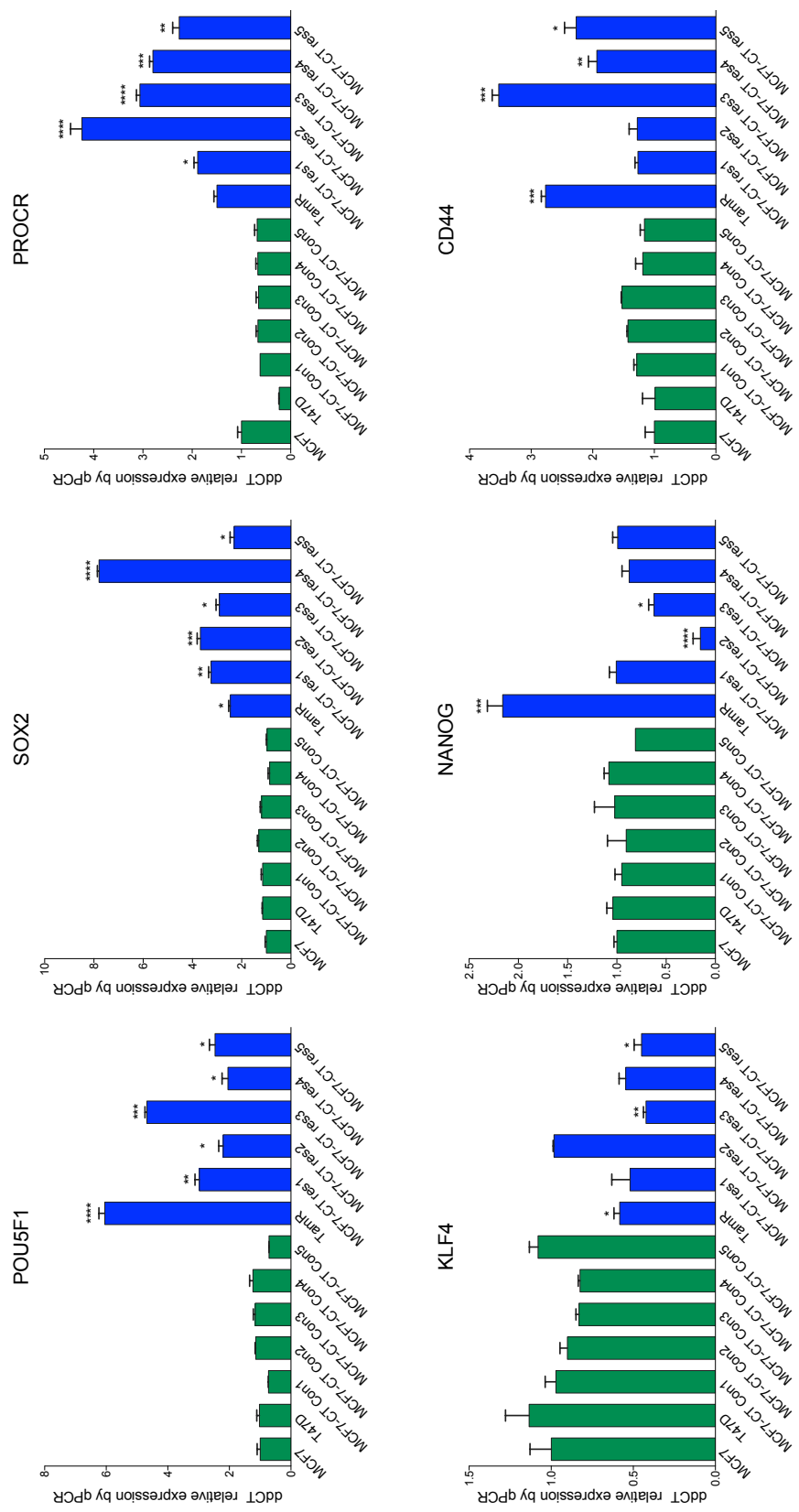
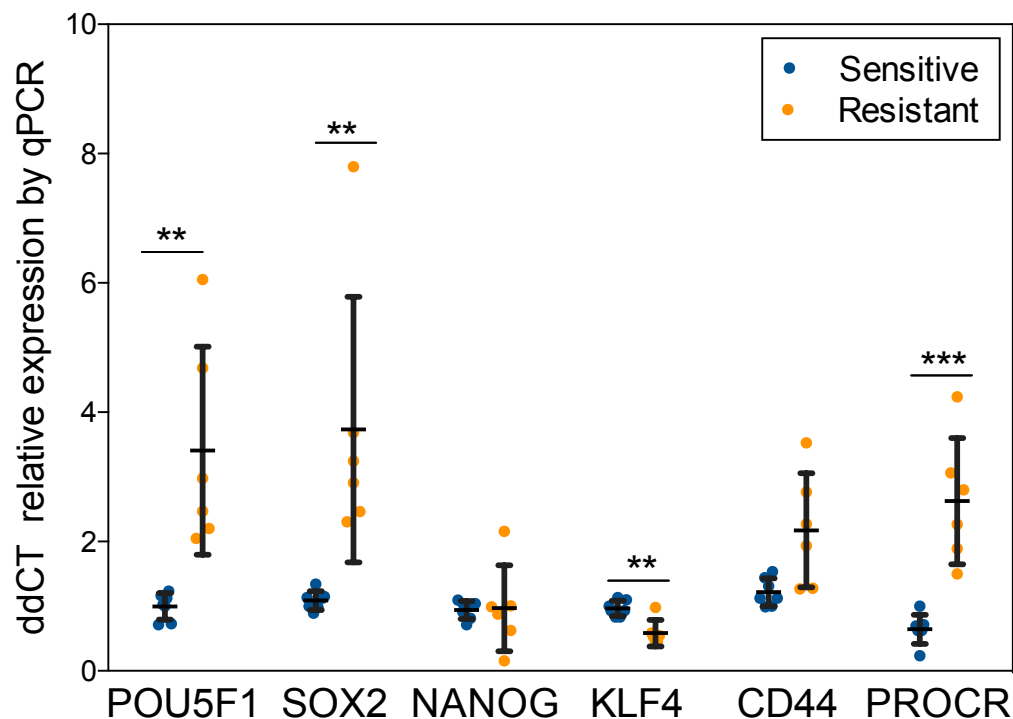


Figure 4.2.2 – Combined analysis of qPCR for stem cell markers**Figure 4.2.2** – Combined analysis of qPCR for stem cell markers

Mean expression values relative to MCF7 from Figure 4.2.1 were combined for resistant and sensitive groups. POU5F1, SOX2, CD44 & PROCR were higher in resistant than sensitive models. KLF4 was low in resistant and NANOG was relatively stable. Statistical significance was computed by one-way ANOVA with Bonferroni correction and is presented versus MCF7 (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$).

Figure 4.2.3 – Mammosphere formation assays

A) 2nd generation mammospheres >50 μm^2 are counted and measured. TamR cell line tends to form significantly smaller mammospheres ($p<0.0001$; two-tailed t-test). **B)** The efficiency of MFE is the percentage of 1st generation mammosphere cells able to self-renew and form new spheres on passage. TamRs have significantly higher MFE than MCF7s, ~1% MFE for MCF7 versus ~2.5% for TamR ($p<0.05$; two-tailed t-test). **C)** MFE was calculated for each MCF7-CT resistant and sensitive lines. Each resistant line had significantly higher MFE than sensitive lines. Statistical significance was computed by one-way ANOVA with Bonferroni correction and is presented versus parental MCF7 (* $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$). Significant heterogeneity exists within the MCF7-CT resistant lines in terms of MFE. MFE is highest in Res5 (~7% MFE) and lowest in Res1 (~2.1%), the difference between these lines is significant ($p<0.001$). **D)** Output of our mammosphere counting workflow in ImageJ. Images are stitched together, and spheres automatically counted and measured. Only those with an area >50 μm^2 are counted. For detailed protocol see Methods, section 2.2.1. **Figure Overleaf.**

Figure 4.2.3 – Mammosphere formation assays

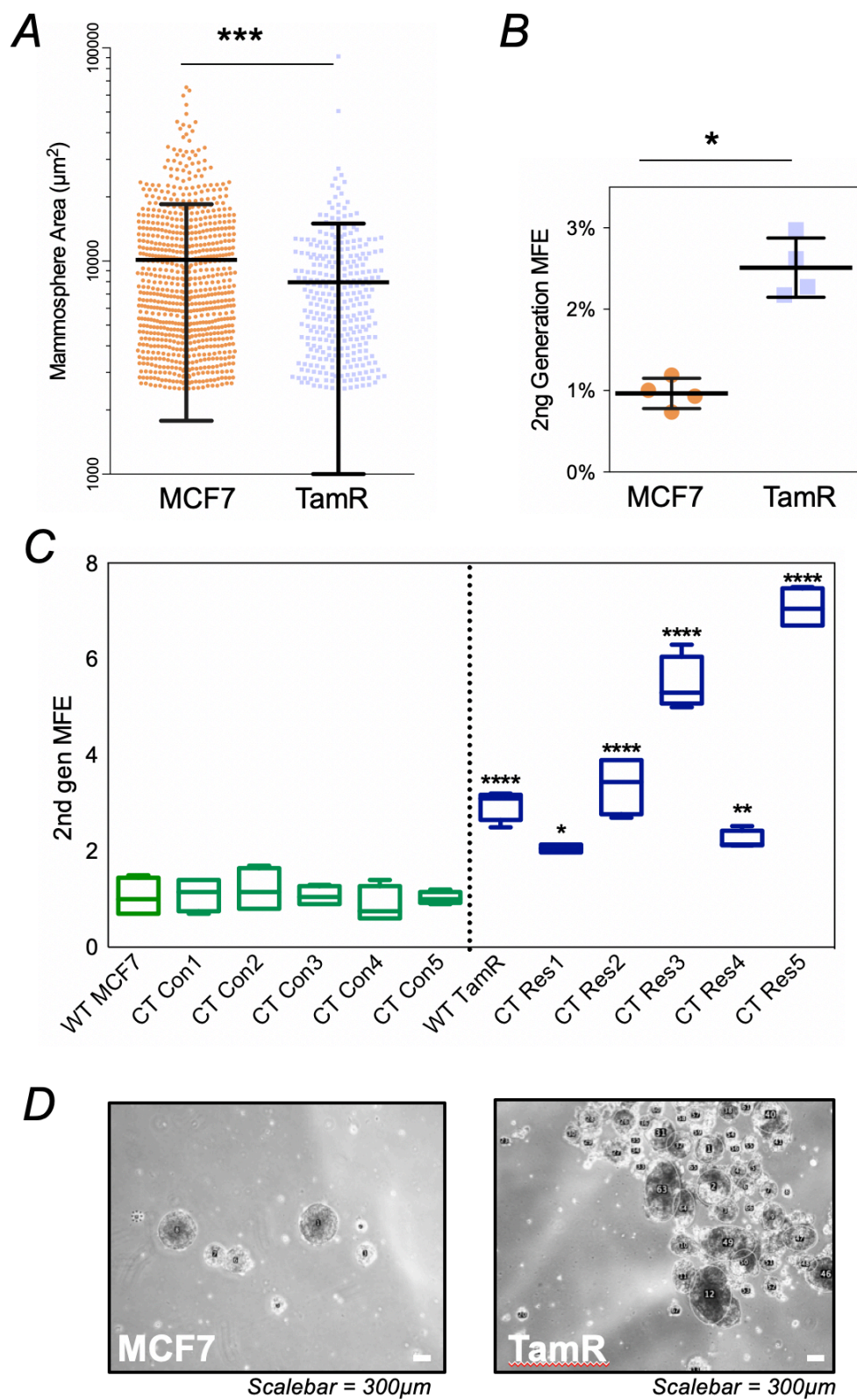


Figure 4.2.4 – Tamoxifen M-IC selection

A) Shows qPCR for *POU5F1*, *SOX2* and *PROCR* in MCF7s and TamRs on treatment with 4-OHT (Tam; 100nM for 24hrs) and/or 10nM E2. *SOX2* and *PROCR* expression was highly significantly reduced in TamRs treated with 4-OHT ($p < 0.001$ in each case). E2 was the only ligand significantly reducing expression of *POU5F1*, though 4-OHT and 4-OHT/E2 combination trended toward this. All three genes were upregulated in MCF7s treated with 4-OHT (alone or in combination, but only statistically significant in 4-OHT only ($p < 0.05$ in each case)). **B)** Mammosphere sizes and MFE for MCF7 and TamR, with and without 4-OHT treatment (100nM 4-OHT or 200nM 4-OHT for 24hrs). 4-OHT treatment increased MFE but decreases size of mammospheres in MCF7s. 4-OHT also significantly reduced the size of TamR mammospheres ($p < 0.001$) but there was no statistically significant effect on TamR MFE. Statistical analysis was by one-way ANOVA with Bonferroni correction for multiple comparisons (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$). **Figure Overleaf.**

Figure 4.2.4 – Tamoxifen M-IC selection

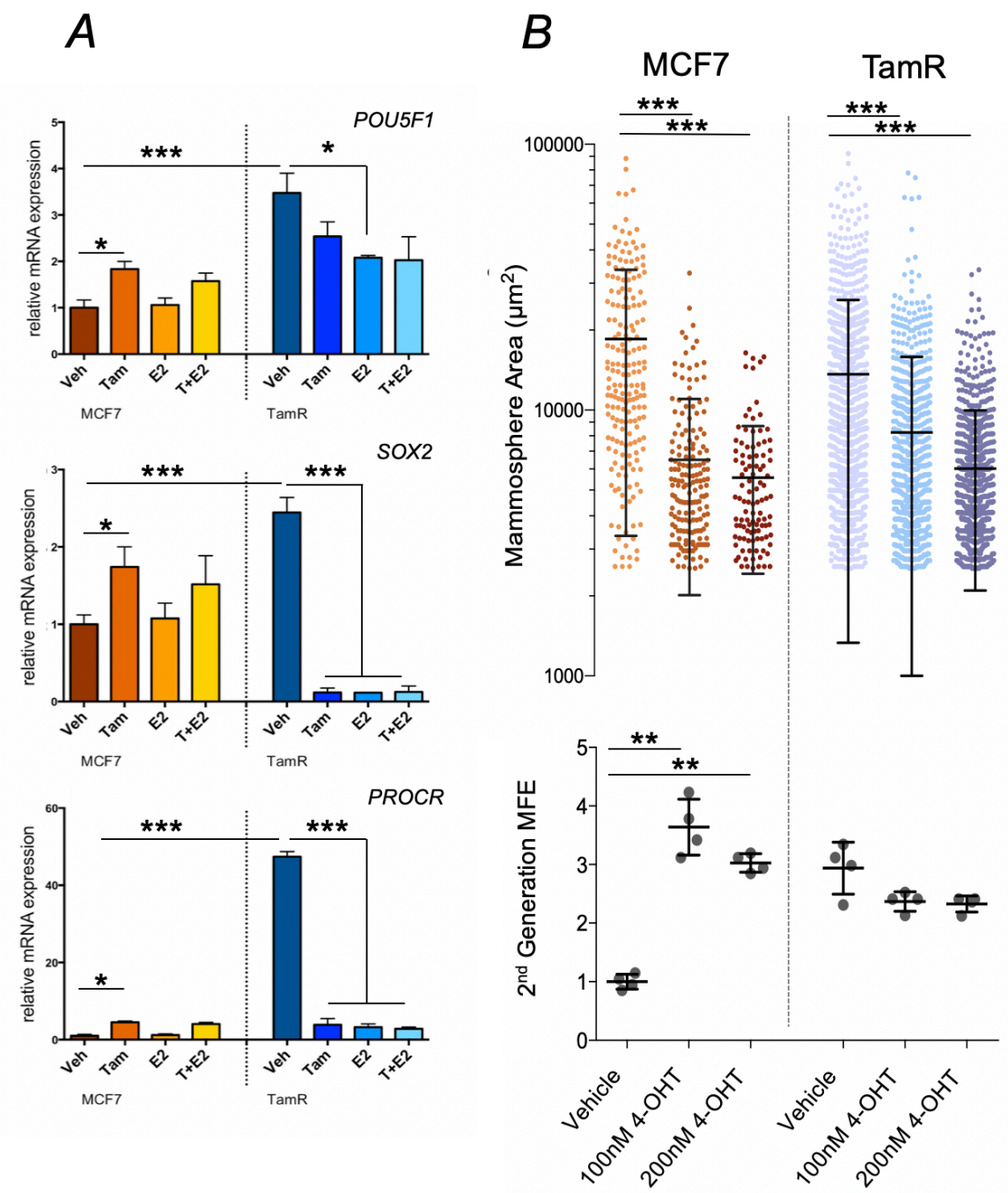


Figure 4.2.5 – OCT4 binding site in ER α promotor

A) Integrative Genomics Browser (IGV) schematic (developed by the Broad Institute; available at software.broadinstitute.org/software/igv/) of the POU5F1 gene region, highlighting, between red lines, a validated promotor site (407) containing ERE binding regions. The region overlaps with both an ER α /FOXA1 binding site previously associated with the development of tamoxifen resistance and poor outcome in ER α + breast cancer patients (248) and a tamoxifen-ER α specific site identified by Welboren et al., (408). **B)** MCF7s treated in 24-well plates with 100nM 4-OHT and harvested periodically over a total of 64hrs. Each timepoint was assayed for gene expression by qPCR (POU5F1 (Oct4), and PGR1 (PR), GREB1 and WISP2; the last three are associated with E2-ER α only binding events (238)). Induction of POU5F1 expression is seen from 24hrs, with corresponding decrease in PGR1, GREB1 and WISP2 expression. This trend is sustained to the end of the assay period. **Figure Overleaf.**

Figure 4.2.5 – OCT4 binding site in ERα promotor

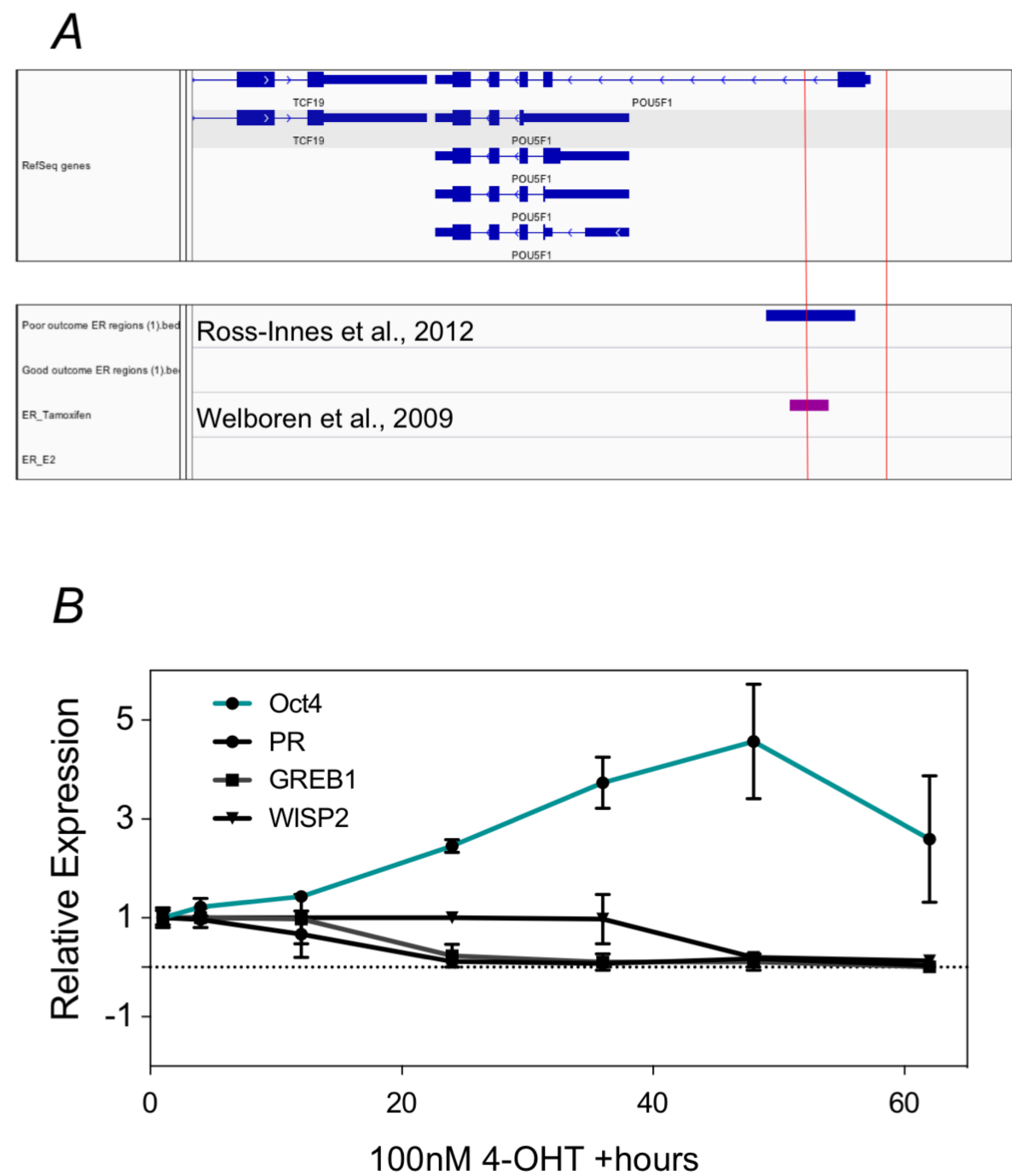


Figure 4.2.6 – Mammospheres are resistant to tamoxifen

A) 2nd generation mammospheres were passaged and plated as monolayers and assayed versus non-mammosphere MCF7 monolayer cultures by dose response for sensitivity to 4-OHT. Mammosphere culture resulted in an IC₅₀ of 0.85µM for MCF7s, more than three times the 0.26µM of monolayer cells ($p < 0.0001$; Extra sum-of-squares F-test). **B)** MCF7s were grown in 8-well CultureSlides pre-coated with 60µl collagen: Matrigel mix and incubated for 2 weeks (for more details see Methods, section 2.2.5). Acinar structures were treated for 24hrs treatment with 0.6µM 4-OHT, resulting in substantial cell death. This result suggests mammosphere cultures, but not other 3D culture techniques, enrich for 4-OHT resistance. **Figure Overleaf.**

Figure 4.2.6 – Mammospheres are resistant to tamoxifen2

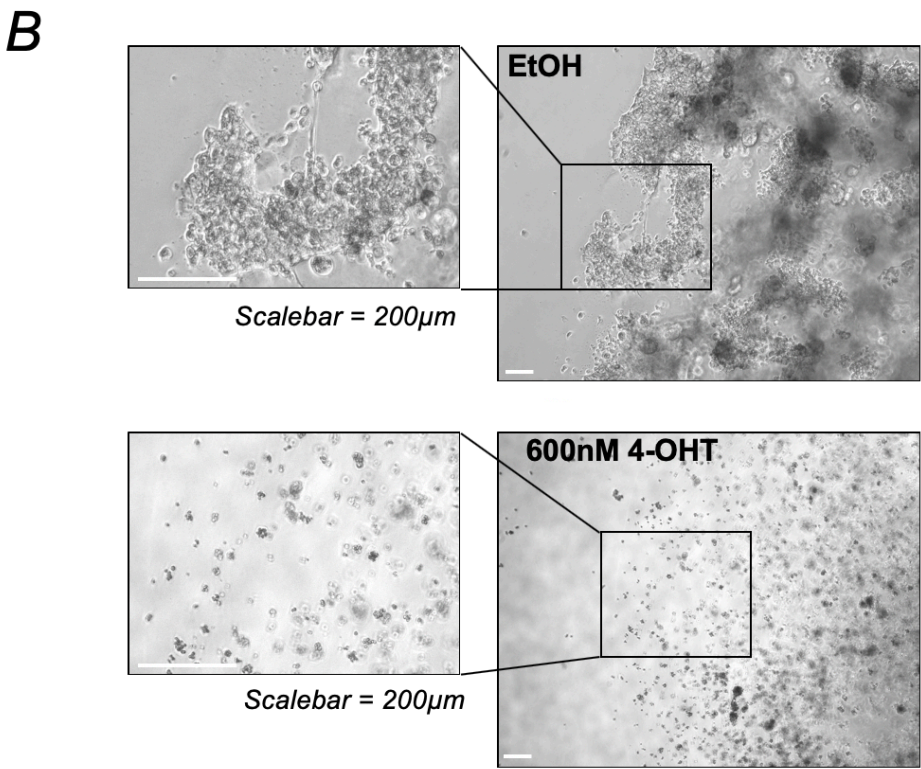
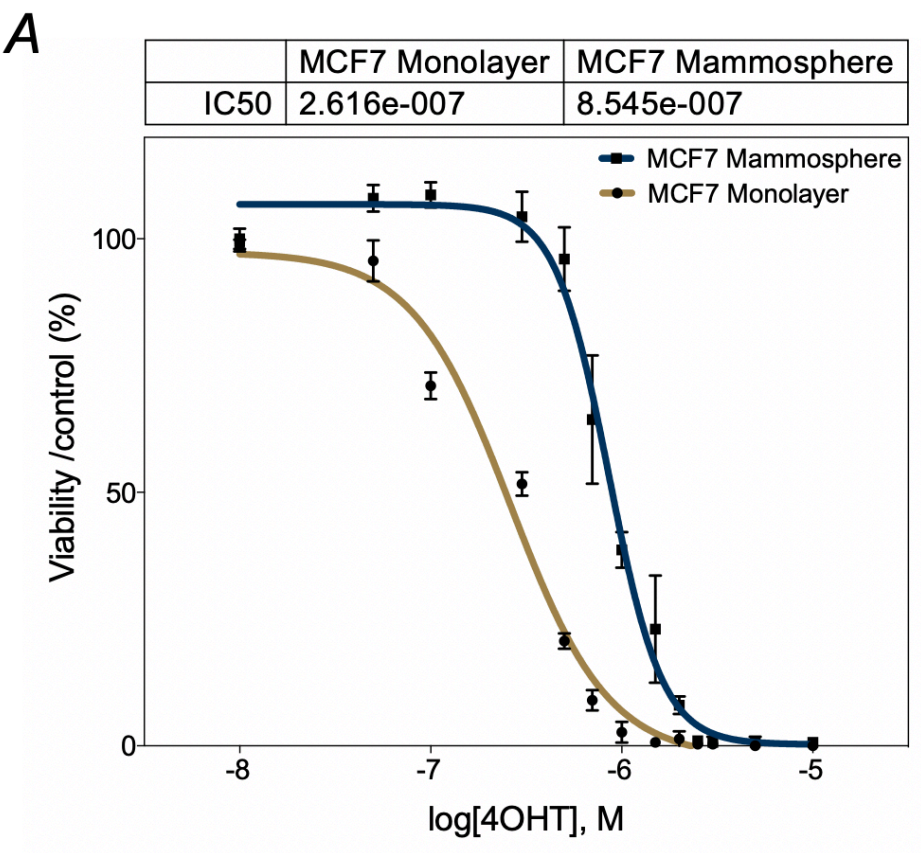
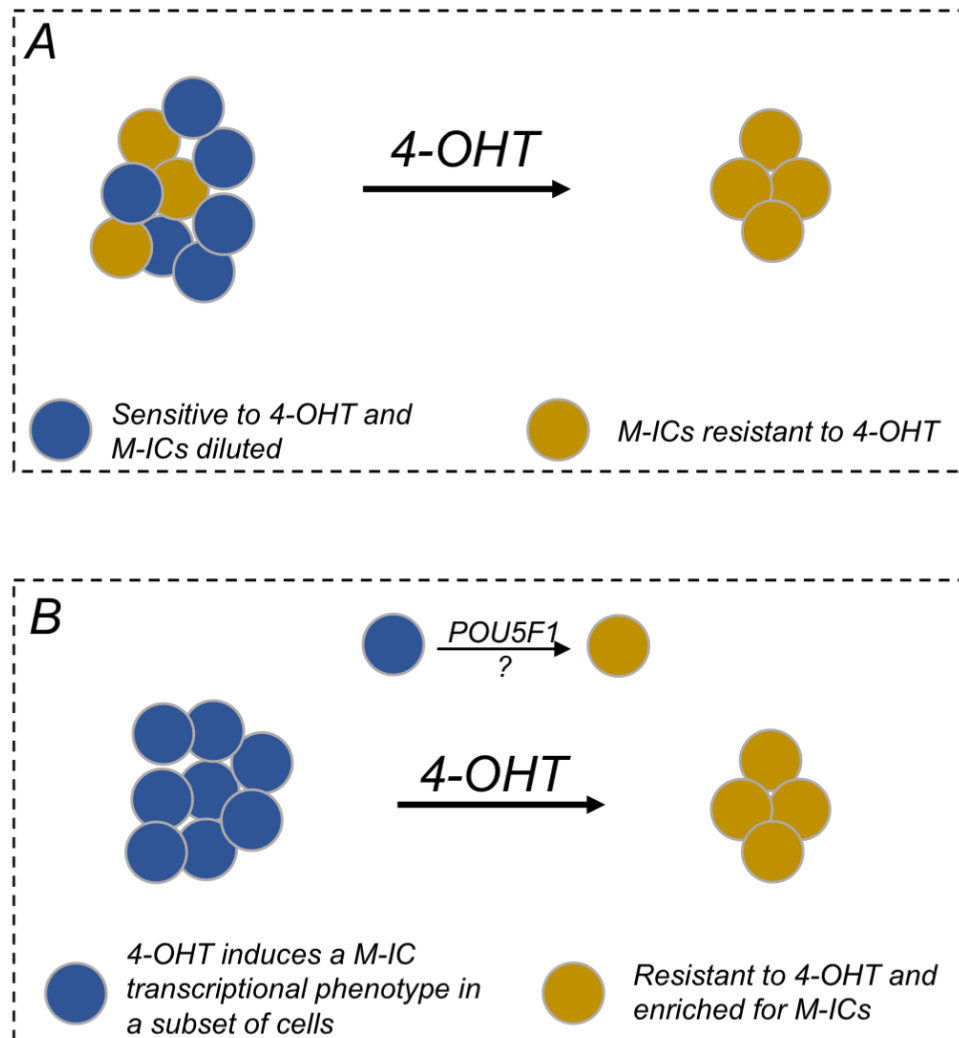
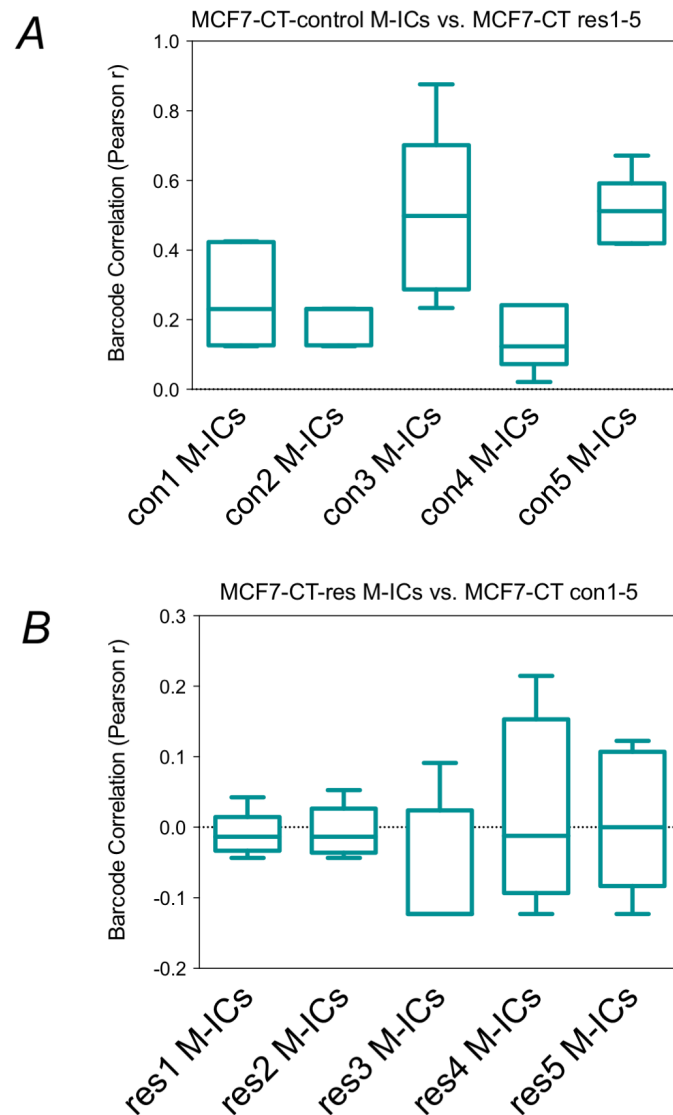


Figure 4.2.7 - Schematic of MCF7 M-IC selection by 4-OHT**Figure 4.2.7** – Schematic of MCF7 M-IC selection by 4-OHT

Schematic depicting two possible routes of M-IC selection by 4-OHT. **A)** A population of cells enhanced in M-ICs and inherently 4-OHT existence is selected for by 4-OHT treatment. **B)** 4-OHT induces a state of plasticity in MCF7s through OCT4/SOX2 etc. which stochastically enables cells to develop resistance.

Figure 4.2.8 – Barcode correlation between M-ICs and 4-OHT resistance**Figure 4.2.8** – Barcode correlation between M-ICs and 4-OHT resistance

A) Correlation between barcodes enriched in M-ICs from each of MCF7-CT control cell lines (con1-5) with each of the 4-OHT resistant cell lines (MCF7-CT res1-5). Data presented as box plot of five pearsons r scores representing the correlation of each MCF7-CT con line with the five resistant lines. **B)** As (A) but MCF7-CT res M-ICs versus each control line (con1-5)

4.3 OCT4 & SOX2 in mammosphere formation

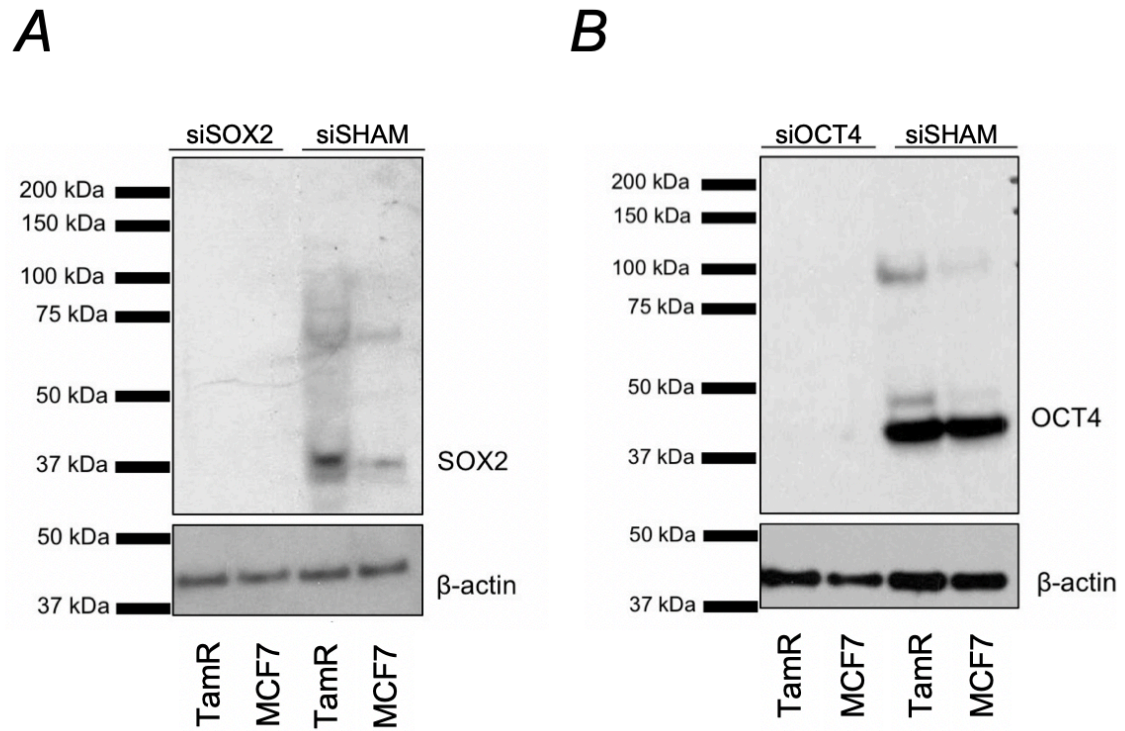
The previous section found that 4-OHT enriches for M-ICs, and for *SOX2* & *POU5F1* expression. These embryonic transcription factors are thought to be responsible for epigenetic activation of cell-type-specific distal regulatory elements involved in determining cell identity during development and differentiation (409,410). Moreover, the ectopic expression of these transcription factors can lead to epigenetic reprogramming and the reacquisition of pluripotency (*i.e.* iPSCs) (410,411). Aside from the key regulatory roles of OCT4, SOX2 and other master transcription factors, this highlights the degree of plasticity retained by the epigenome (401,411,412). Cancer cells have also been shown to remodel the epigenomic landscape in response to the selective pressures associated with therapy (86,413), for example, in the adaptation to oestrogen deprivation (413) and to treatment with AIs (401), and there is a specific association between the CSC phenotype and epigenetic remodelling (414). In this section, we sought to directly test the hypothesis that transcription factors SOX2 and OCT4 contribute to mammosphere formation (as an *in vitro* model of BC-TICs) and/or resistance to 4-OHT. 1st generation mammospheres were generated either in the presence of sham siRNAs (for GFP) or targeted siRNAs for *SOX2/POU5F1* with/without exposure to 100nM 4-OHT for 24 hrs. 2nd generation mammospheres were then generated and MFE scored based on viable input cell counts and spheres formed in 2nd generation culture, after four days. Efficient siRNA knockdown was confirmed by western blot (*Figure 4.3.1 A & B*).

Treating MCF7s with 4-OHT appeared to have little effect on gross sphere morphology (relative to Sham controls), whereas siOCT4 visually appeared to reduce sphere size and potentially aggregation (*Figure 4.3.2 A*). As before, treatment with 4-OHT reduced the size of mammospheres formed but enriched for M-ICs (*Figure 4.3.2 B*), thus enhancing MFE. Treatment with siOCT4 reduced

mammosphere growth (not significant) and slightly increased M-IC number (not significant), perhaps suggesting that aggregation of smaller spheres into larger was inhibited. Interestingly, 4-OHT treatment in the presence of siOCT4 completely ablated the M-IC enhancing effects of 4-OHT ($p<0.01$; one-way ANOVA with Bonferroni correction). Furthermore, MFE was significantly reduced by 4-OHT in the presence of siOCT4 relative to siOCT4 alone ($p<0.05$; one-way ANOVA with Bonferroni correction), suggesting that 4-OHT becomes toxic to MCF7 mammospheres in the absence of OCT4. In contrast, siSOX2 almost completely removed the ability of cells to form mammospheres in either the presence or absence of 4-OHT, suggesting that SOX2 is a requirement for cells to form mammospheres (*Figure 4.3.2 A and B*). This appears to be in agreement with the literature, with several studies finding SOX2 is activated in breast CSCs/BC-TICs (415) and that it may be essential to mammosphere formation via its downstream target Mucin-1 (416).

Our studies have shown that resistant lines tend to have increased basal expression of *SOX2* and *POU5F1*, but that these tend to be downregulated on 4-OHT treatment. We next asked whether these transcription factors contributed to survival of the M-IC fraction of TamR cells. *Figure 4.3.3 A & B* show that both siOCT4 and siSOX2 drastically reduced mammosphere formation in TamRs. As before (*Figure 4.3.2 B*) we see a slight, non-significant decline in the MFE of TamRs after treating with 4-OHT (*Figure 4.3.3 A & B*). *POU5F1*/OCT4 appears to have only a slight role in untreated MCF7 M-IC number but is completely required for mammosphere formation in TamRs.

Together, these results highlight the importance of OCT4 and SOX2 in the formation and/or maintenance of M-ICs and 4-OHT resistance. Interestingly, OCT4 appears to be specifically important in the 4-OHT induced M-IC enrichment process, in MCF7s.

Figure 4.3.1 –SOX2/OCT4 knockdowns**Figure 4.3.1** – SOX2/OCT4 knockdowns

A) Western blot for SOX2 in MCF7s and TamR cells with and without siSOX2. Efficient reduction in SOX2 protein is seen. **B)** Western blot for OCT4 in MCF7s and TamR cells with and without siSOX2. Efficient reduction in OCT4 protein is seen.

Figure 4.3.2 – Role of SOX2/OCT4 in MCF7 MFE

A) Representative phase contrast microscopy of mammospheres formed by MCF7s with/without 4-OHT pretreatment (200nM for 24hrs) and/or siRNA for OCT4/SOX2. Mammospheres appear smaller in siOCT4 and siSOX2 without 4-OHT. With 4-OHT, both siOCT4 and siSOX2 show very few mammospheres. **B)** Quantification of (A) showing a significant increase in MFE with 4-OHT and Sham siRNA ($p<0.01$). This was completely ablated by the addition of siOCT4, suggesting an essential role for OCT4 in the formation of mammospheres and/or enrichment of M-ICs by 4-OHT in MCF7s. SOX2 siRNA substantially reduced the number of mammospheres formed irrespective of 4-OHT treatment, suggesting an essential role in M-IC maintenance. Statistical analysis was by one-way ANOVA with Bonferroni to correct for multiple comparisons (* $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$). **Figure Overleaf.**

Figure 4.3.2 – Role of SOX2/OCT4 in MCF7 MFE

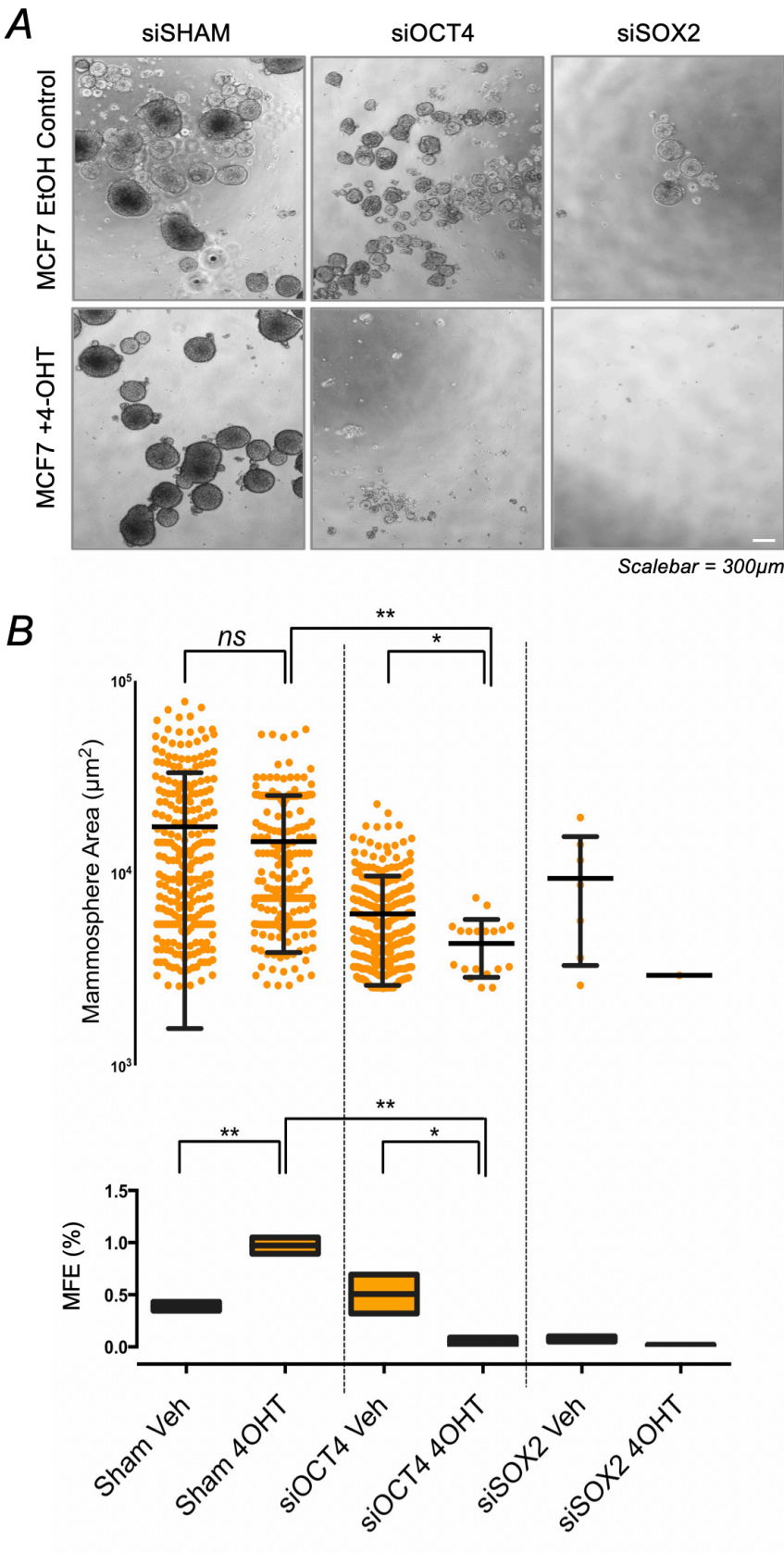
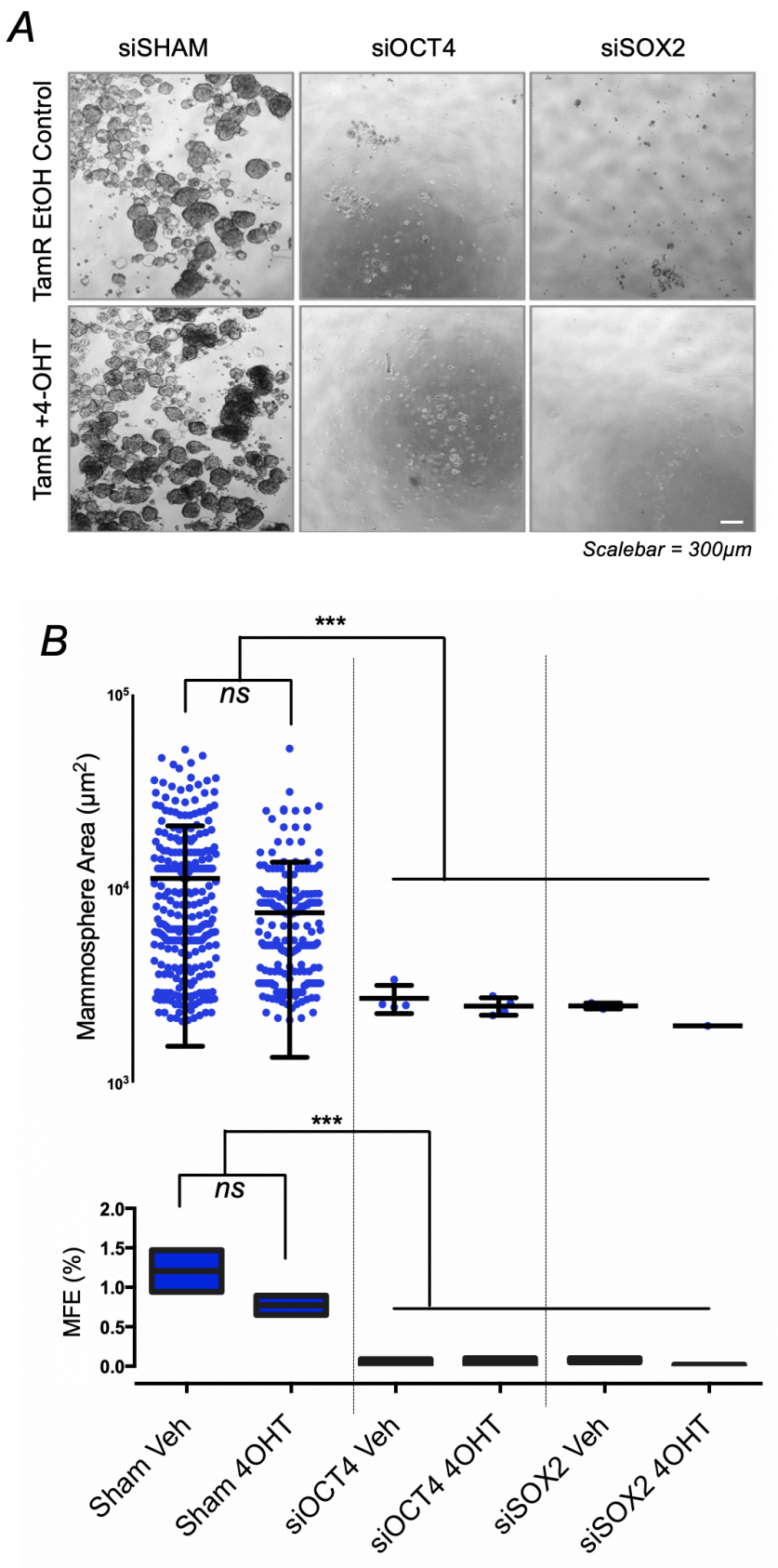


Figure 4.3.3 – Role of SOX2/OCT4 in TamR MFE

A) Representative phase contrast microscopy of mammospheres formed by TamRs with/without 4-OHT pretreatment (200nM for 24hrs) and/or siRNA for OCT4/SOX2. Mammospheres failed to form when either siRNA was used. **B)** Quantification of (A) showing no significant differences in size or MFE of TamRs with the addition of 4-OHT and a highly statistically significant drop in mammosphere size and number when either siOCT4 or siSOX2 was used ($p < 0.001$). Statistical analysis was by one-way ANOVA with Bonferroni to correct for multiple comparisons. **Figure Overleaf.**

Figure 4.3.3 – Role of SOX2/OCT4 in TamR MFE



4.4 Generation of OCT4-Vex-Poly MCF7 Cell Line

In order to better study the role of OCT4 in the development of tamoxifen resistance and in the regulation of M-ICs, we generated an MCF7 OCT4-vexGFP *POU5F1* over expressing cell line. The expression insert can be seen in *Figure 4.4.1 A*, *POU5F1* is downstream of a phosphoglycerate kinase (PGK) promoter (329). Parental MCF7s were infected with lentiviral constructs containing OCT4-vexGFP, confirmation of efficient infection by flow cytometry is visualised in *Figures 4.4.1 B & C*. With overall positivity of 83.7% GFP, we opted to sort a selection of medium-intensity GFP-positive cells, in order to capture some population heterogeneity in terms of viral integrations and *POU5F1* expression. The resulting cell line, OCT4-Vex-Poly was maintained in standard monolayer culture conditions used for parental MCF7s. *Figure 4.4.1 D* shows qPCR results for *POU5F1* expression, showing significantly greater mRNA representation of *POU5F1* in the OCT4-Vex-Poly line relative to parental and TamR.

We next characterised the OCT4-Vex-Poly cell line for *in vitro* proliferation and 4-OHT sensitivity. *Figure 4.4.2 A* shows significantly slower proliferation in OCT4-Vex-Poly cells versus parental ($p=0.0432$, non-linear mixed effect model). A key feature of tamoxifen resistant lines in our studies has been quiescence or slow cell cycling, we reasoned that had OCT4-Vex-Poly taken on drug resistant phenotype, we would observe slower growth. *Figure 4.4.2 B* shows the result of dose response calculations for OCT4-Vex-Poly, which is slightly more resistant to 4-OHT than parental (IC₅₀'s of 0.68 μ M and 0.27 μ M respectively; $p<0.05$ by Extra sum-of-squares F-test).

We next sought to verify reports in the literature linking OCT4 to BC-TIC-like activity through the *in vitro* mammosphere formation surrogate assay. We calculated MFE for OCT4-Vex-Poly, MCF7 parental and MCF7 parental plus OCT4 siRNA, with each 1st generation mammosphere culture treated with EtOH vehicle or 200nM 4-OHT

(Figure 4.4.3 A). Surprisingly, OCT4-Vex-Poly had no greater ability to form 2nd generation mammospheres than MCF7 parental (plus a control siRNA for GFP). However, we confirmed earlier findings that pre-treatment with 4-OHT significantly increased MFE in the parental MCF7 cell line ($p < 0.01$). Moreover, we found this affect was markedly amplified in the OCT4-Vex-Poly cell line. OCT4-Vex-Poly had 10-fold greater MFE when pre-treated with 4-OHT compared to control ($p < 0.001$). As we had found earlier, siRNAs for *POU5F1* made little difference to M-ICs when cells were treated with EtOH alone, but with the addition of 4-OHT, MFE was completely ablated ($p < 0.05$; all by one-way ANOVA with Bonferroni correction). These results are shown visually by representative bright field images in Figure 4.4.3 B.

Mammosphere assays have been shown to predict BC-TIC content by measuring M-IC numbers in an *in vitro* surrogate of a tumour formation assay. However, the assay has many well-known limitations, not least that aggregation of smaller spheres into larger ones could impact on assay sensitivity. We sought to confirm the synergistic role of OCT4 and 4-OHT treatment in enriching for M-ICs, and so performed *in vitro* limiting dilution assays (LDA; Figure 4.4.4). Based on seeding increasingly small numbers of single cells generated from 1st generation mammospheres, and counting the resulting spheroids formed. We employed the ELDA statistical package (324) to estimate the functional M-IC content of each group. In the untreated control, OCT4-Vex-Poly had the highest M-ICs frequency at 1/51.5 cells, this was 1/99.4 for MCF7 parental (plus control siRNA) and 1/872.7 for siOCT4 (Figure 4.4.4 A). When we dose 1st generation mammospheres with 4-OHT prior to conducting the assay, OCT4-Vex-Poly had the highest M-IC frequency at 1/13.2 cells, this was 1/36.2 for MCF7 parental (plus control siRNA) and 1/1452.2 for siOCT4 (Figure 4.4.4 B). In both the control and treated groups each cell population was highly statistically significantly (p value reported is calculated via χ^2) different from each of the other

two groups, giving us a high confidence that, at least *in vitro*, there is a true effect of OCT4 on increasing M-IC number.

Figure 4.4.1 – Generation of OCT4-Vex-Poly MCF7 cell lines

A) Diagram of OCT4-Vex-GFP insert. Insert was packaged in lentiviral vector (2nd generation) as described in Methods, section 2.3.2, and lentiviral particles used to infect MCF7s. **B)** Shows flow cytometry-based analysis of infection efficiency, mode normalised GFP detection by 405nm laser is shown. **C)** Shows derivation of the OCT4-Vex-Poly cell line from the infected MCF7 population. OCT4-Vex-Poly were cell sorted from the central region of VexGFP positive cells. 83.7% of cells were positive for vexGFP. **D)** Confirmation by qPCR of over-expression of POU5F1 in OCT4-Vex-Poly cell line versus TamR and MCF7 ($p < 0.001$). OCT4-Vex-Poly shows 4-fold overexpression relative to TamR and 7-fold relative to MCF7.

Figure Overleaf.

Figure 4.4.1– Generation of OCT4-Vex-Poly MCF7 cell lines

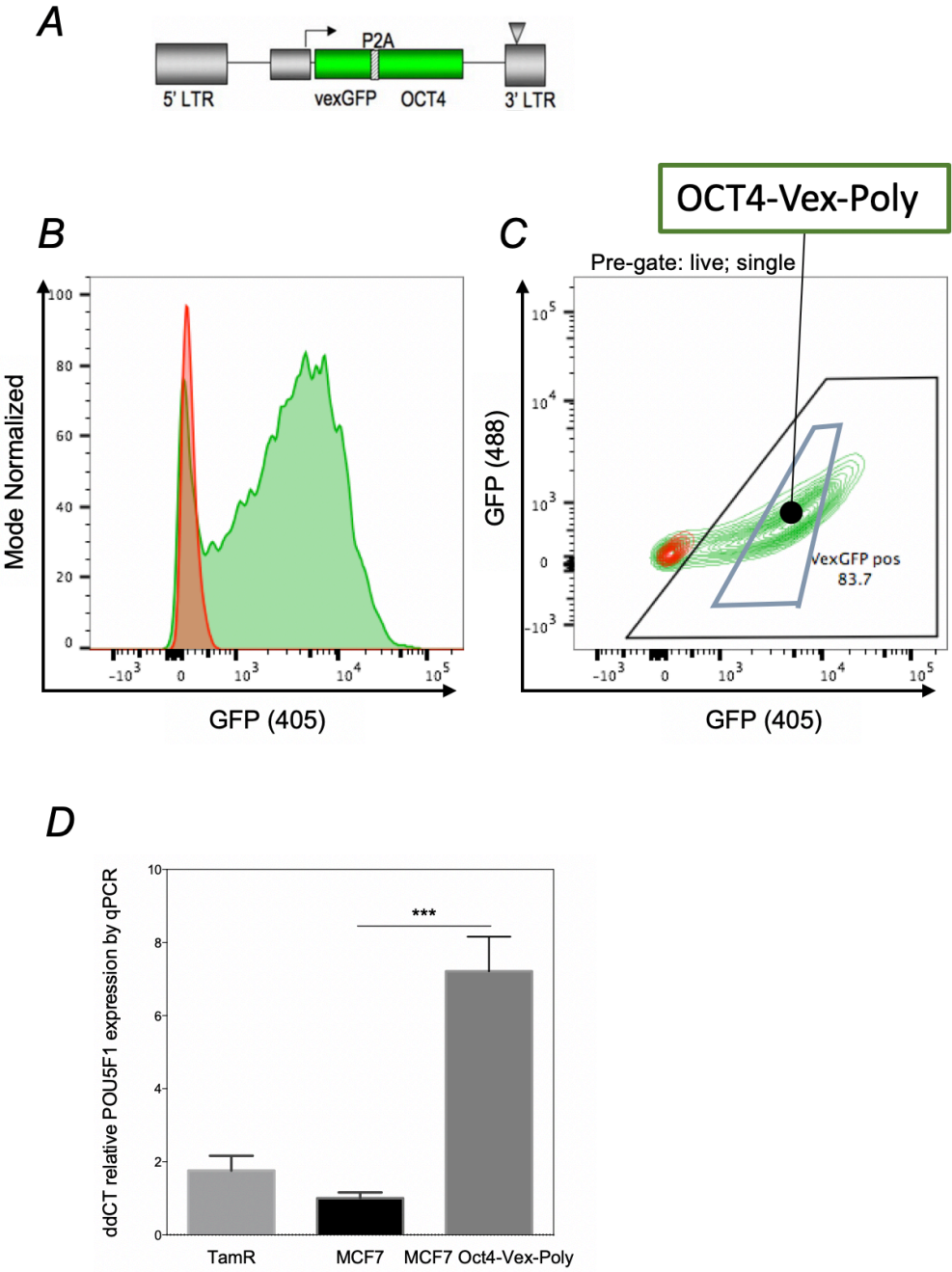


Figure 4.4.2 – Characterization of OCT4-Vex-Poly MCF7s

A) *In vitro* proliferation of OCT4-Vex-Poly versus MCF7 parental in the IncuCyte live cell imaging system. OCT4-Vex-Poly was significantly slower growing than parental ($p < 0.05$ by non-linear mixed effects model). **B)** Shows dose response analysis for OCT4-Vex-Poly to calculate 4-OHT sensitivity. OCT4-Vex-Poly is slightly more resistant to 4-OHT than parental (IC₅₀'s of 0.68 μ M and 0.27 μ M respectively; $p < 0.05$ by Extra sum-of-squares F-test). **Figure Overleaf.**

Figure 4.4.2 – Characterization of OCT4-Vex-Poly MCF7s

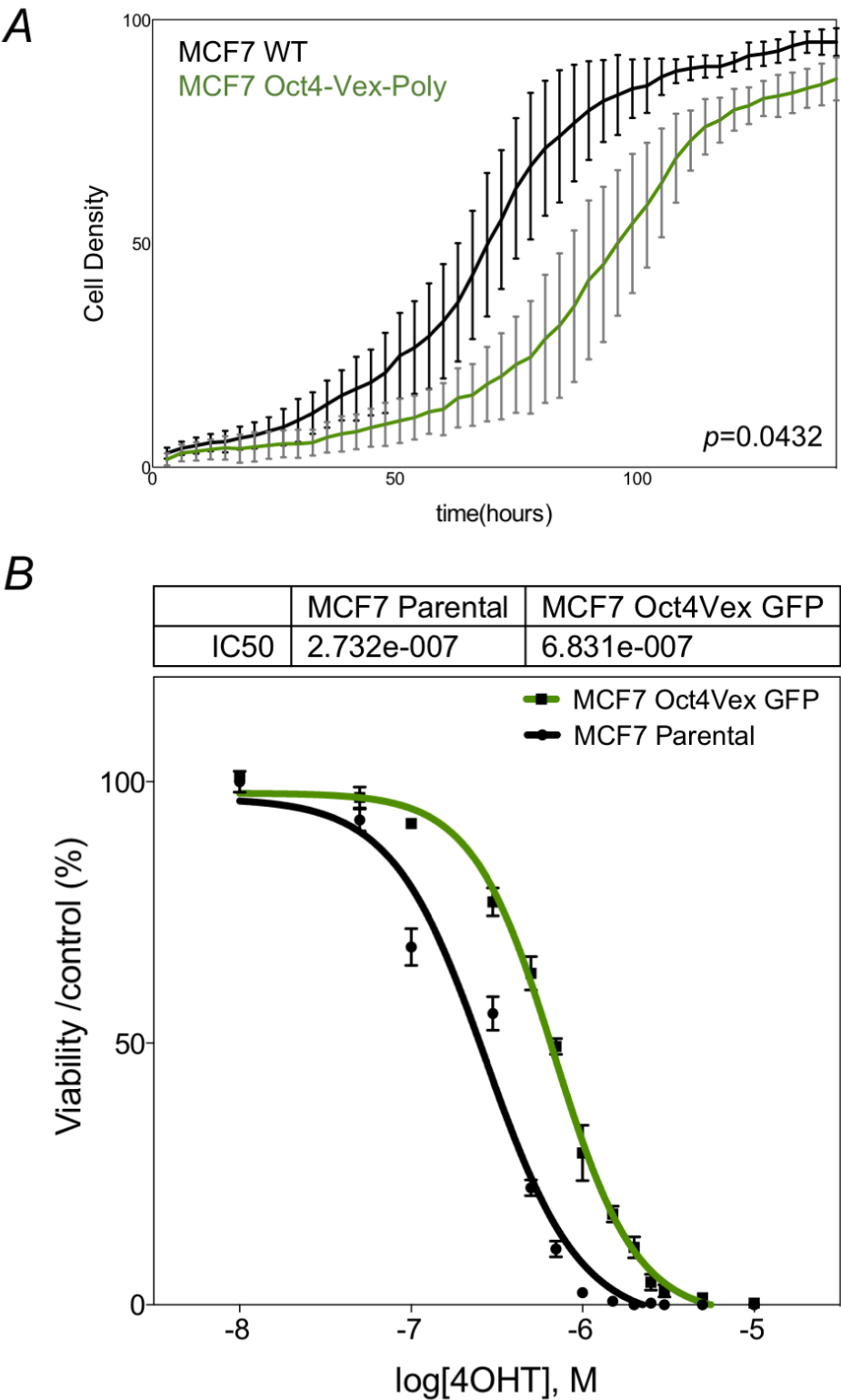


Figure 4.4.3 – MFE in OCT4-Vex-Poly MCF7s

A) Calculated MFE for OCT4-Vex-Poly, MCF7 parental and MCF7 parental plus OCT4 siRNA cell lines (+/- 4-OHT pre-treatment at 300nM 48hr). Several conclusions can be drawn: OCT4-Vex-Poly had no greater ability to form 2nd generation mammospheres than MCF7 parental; pre-treatment with 4-OHT significantly increased MFE in the parental MCF7 cell line ($p < 0.01$); this affect was amplified in the OCT4-Vex-Poly cell line (10-fold greater MFE when pre-treated with 4-OHT ($p < 0.001$)) & siOct4 made little difference to vehicle treated MFE, but completely ablated MFE with 4-OHT treatment ($p < 0.05$; all by one-way ANOVA with Bonferroni correction). **B)** Representative images from the analysis in (A). **Figure Overleaf.**

Figure 4.4.3 – MFE in OCT4-Vex-Poly MCF7s

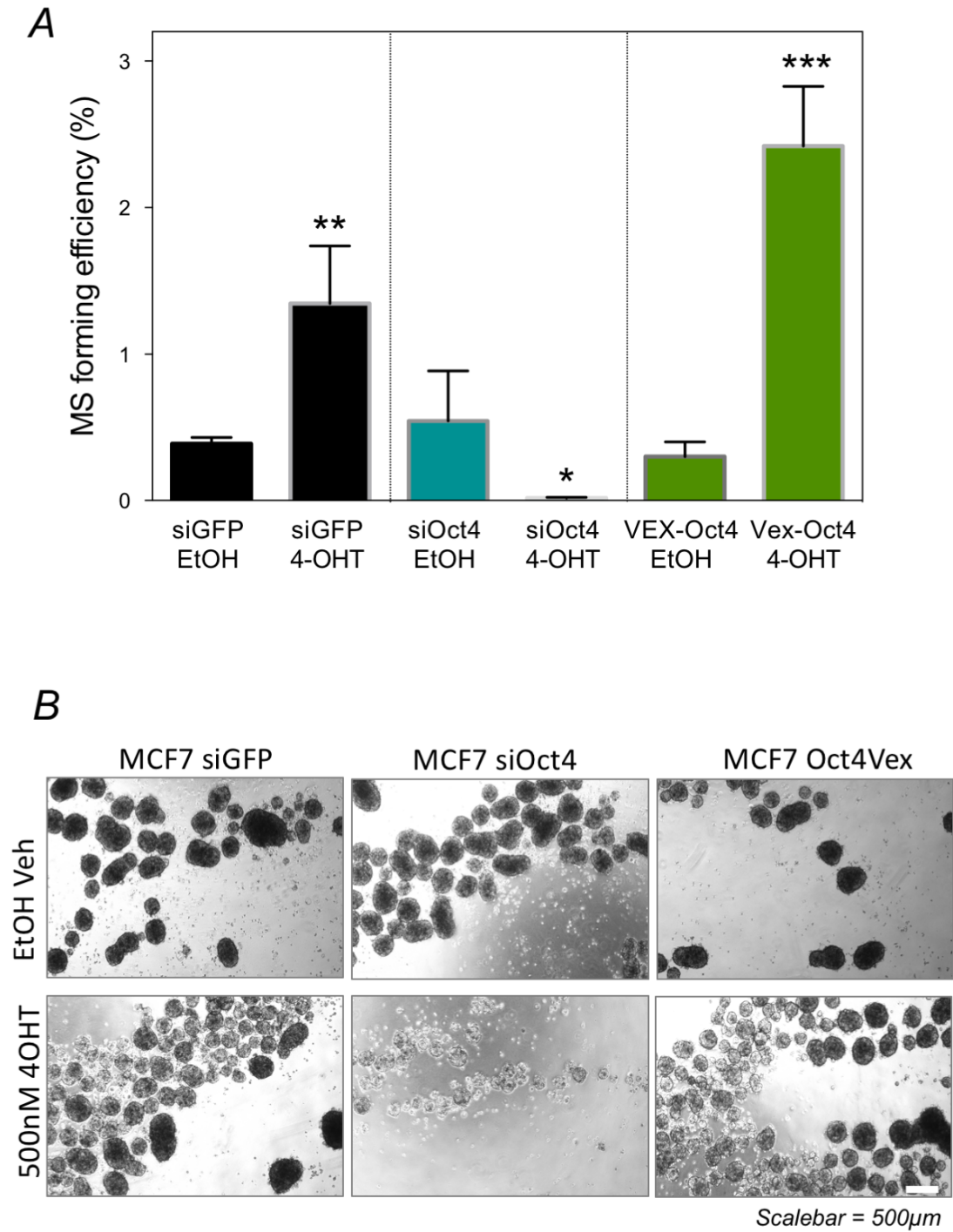
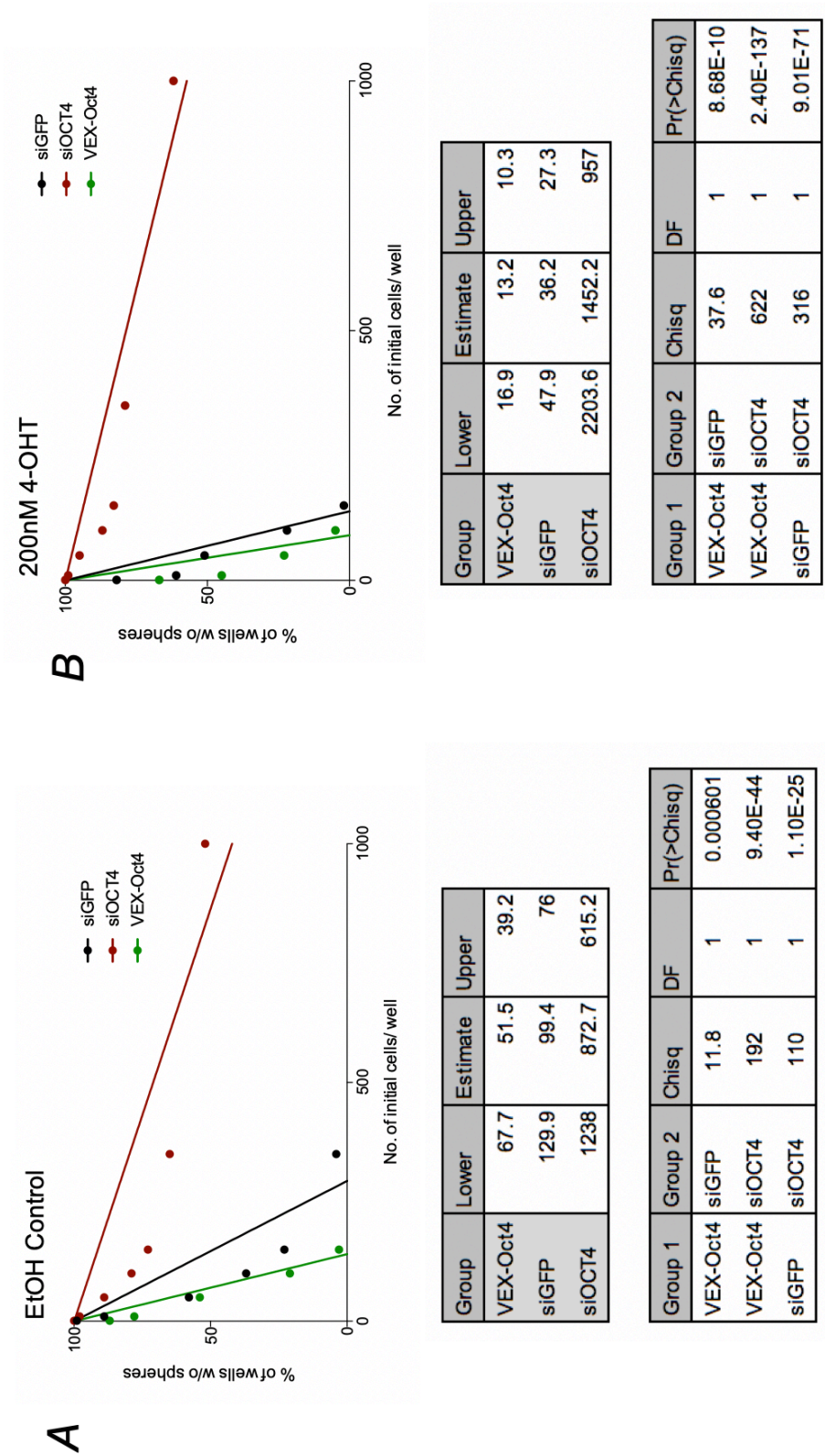


Figure 4.4.4 – *in vitro* LDA of OCT4-Vex-Poly MCF7s

A) Increasingly small numbers of single cells generated from 1st generation mammospheres were seeded in mammosphere conditions, and spheroids counted after 3 weeks in culture. Further details on seeded cell numbers can be found in Methods, section 2.2.6. The ELDA statistical package (324) was employed to estimate functional M-IC cell count in each group. Output of this analysis is shown in the accompanying table: OCT4-Vex-Poly had the highest M-IC frequency at 1/51.5 cells, MCF7 parental (plus control siRNA) was 1/99.4 and siOCT4 was 1/872.7. Each cell population was highly statistically significantly different from each of the other two groups (*p* value reported is calculated via χ^2). **B)** As (A), but cells dosed with 200nM 4-OHT 48hrs during 1st generation mammosphere growth. Here, OCT4-Vex-Poly had the highest M-IC frequency at 1/13.2 cells, MCF7 parental (plus control siRNA) had 1/36.2 and siOCT4 had 1/1452.2. As before, each cell population was highly statistically significantly different from each of the other two groups (*p* value reported is calculated via χ^2). **Figure Overleaf.**

Figure 4.4.4 – in vitro LDA of OCT4-Vex-Poly MCF7s



4.5 EMT in OCT4-Vex-Poly, TamR and MCF7s

Numerous studies have linked breast cancer drug resistance to the process of EMT (104,393), and EMT to the acquisition of a BC-TIC-like phenotype (394,400). With one study, for example, linking *DMXL2* expression to Notch pathway activation and EMT in ER α positive breast cancer patients that progress after endocrine therapy (366). Notch itself has been shown to regulate cell fate (417), proliferation (418), apoptosis (419), as well as stem cell survival and self-renewal (417,420,421). In breast cancer, increased expression of *Notch-1* predicts poorer overall survival (422) and has been shown to promote proliferation and CSC survival in a subset of HER2+ disease (423).

We next sought to understand the potential contributions of EMT to tamoxifen resistance in our models. Mammospheres from OCT4-Vex-Poly experiments (*Figure 4.4.3*) were harvested and prepared for IHC (*Figure 4.5.1*). ER α and Ki67 were seen to colocalise in OCT4-Vex-Poly, suggesting at least some oestrogen dependence. Moreover, OCT4-Vex-Poly mammospheres appeared variably enriched for the ALDH1A marker of CSCs/BC-TICs. N-Cadherin, E-Cadherin and EpCam were also profiled to investigate whether OCT4-Vex-Poly mammospheres had undergone EMT. However, no major differences between markers expressed by MCF7 parental and OCT4-Vex-Poly mammospheres is immediately obvious. As OCT4-Vex-Poly are enriched for M-ICs resistance to 4-OHT, it is interesting that they do not share properties of EMT.

We sought to further profile EMT in 4-OHT resistant cell lines. TamR cells tend to have a more mesenchymal stellate phenotype when grown as a monolayer, *Figure 4.5.2 A* (393). Their migration, measured by a wound healing assay, is also significantly faster than parental MCF7s, though both cell lines are inhibited in this regard by 4-OHT, suggesting a functional role for ER α signalling and migration even

in resistant models, *Figure 4.5.2 B & C*. Future studies should in, for example, MDA-MB-231 cells as a positive control for a migratory phenotype. EMT is associated with differential expression of key marker genes. Putative makers of an epithelial type phenotype include E-Cadherin, Claudin3 and Claudin4 whereas markers of a more mesenchymal phenotype include SNAI1, TGFB1, Vimentin and Twist (209,424). By comparing gene expression patterns of TamRs and MCF7s in monolayer and mammosphere culture, we find that mesenchymal-type genes dominate in TamRs whereas epithelial-type genes dominate in MCF7s (*Figure 4.5.3 A*). This is most apparent in mammosphere (3D) conditions.

We sought to confirm these results by immunofluorescence in TamR and MCF7 monolayers (*Figure 4.5.3 B*). Levels of Vimentin and E-Cadherin appear to correlate with mRNA levels from *Figure 4.5.3 A*, though the expected mesenchymal phenotype of cells having undergone EMT is not perfectly observed. For example, Vimentin is expected to be up-regulated significantly after EMT and follow a fibrous type expression pattern, whereas in our studies expression changes are minor and Vimentin appears to be localised in foci in TamRs (425). Vimentin intermediate filaments have been shown to interact with the centrosome if expression is low (426), suggesting that whilst Vimentin is up-regulated in TamRs protein levels are still low. Interestingly, E-Cadherin loss is seen as the hallmark of EMT and TamR cells show evidence of two distinct populations distinguished by differential E-Cadherin expression. One population expresses E-Cadherin and SNAI1 similar in pattern to parental MCF7s, the other expresses Vimentin foci along with SNAI1.

As mammospheres are enriched for mesenchymal markers by qPCR, these results indicate that the E-Cadherin^{low} subpopulation of TamR cells may account for its increased M-IC number. However, E-Cadherin has been shown to be essential in the formation of mammospheres. *Iglesias et al.*, found that knock down of E-cadherin in MCF7s completely ablated their ability to form mammospheres (427). The authors also found that whilst SKBR3 cells could not normally sustain long term

mammosphere cultures, over-expression of E-cadherin allowed them to form mammospheres (427). It is therefore unclear whether the presence of an E-Cadherin^{low} subpopulation, enriched in EMT markers, could mark M-ICs the TamR cell line.

Figure 4.5.1 – Mammosphere IHC

Mammospheres from Figure 4.5.5.1 were harvested and prepared for IHC. ER α and Ki67 were seen to colocalise in OCT4-Vex-Poly. OCT4-Vex-Poly mammospheres appeared enriched for ALDH versus parental MCF7s. No major differences between EMT markers (N-Cadherin, E-Cadherin and Epcam) was obvious, suggesting that whilst this line has similar properties to BC-TIC-like-cells, it has not undergone EMT.

Figure Overleaf.

Figure 4.5.1 – Mammosphere IHC

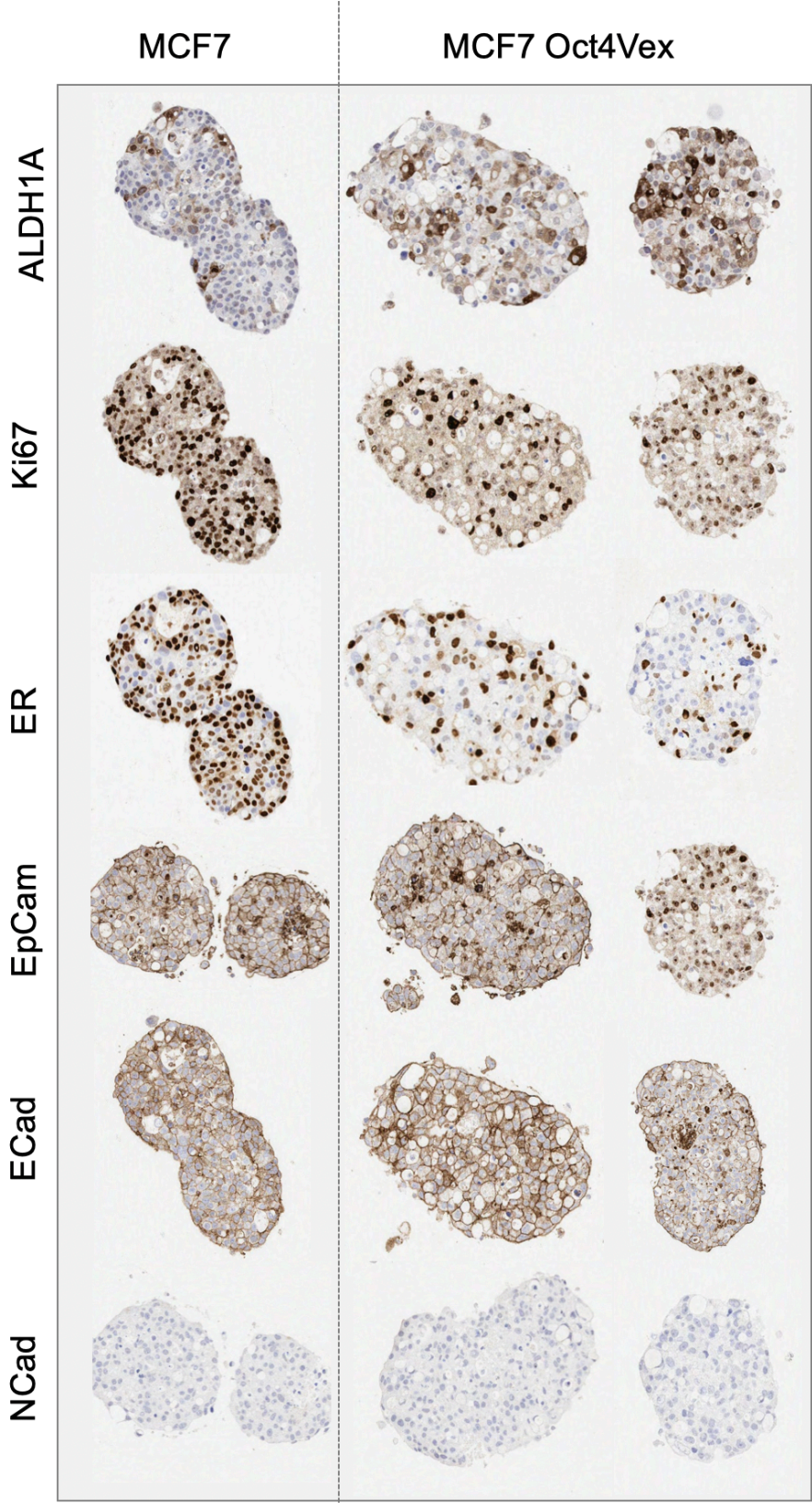


Figure 4.5.2 – TamR cells have a migratory phenotype

A) Phase contrast microscopy of MCF7 and TamR cells, showing stellate and slightly mesenchymal-like TamR cell shape. **B)** Wound healing assay showing TamRs are significantly ($p < 0.001$) more migratory than MCF7s. In both cell lines, 4-OHT is inhibitory to migration at 100nM, suggesting ER α response is still active in TamRs in terms of migratory repression. Statistical analysis was by non-linear mixed effect model (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$). **C)** Phase contrast of wound healing assay as carried out using the Essen Biosciences 96 well wound maker and associated software, images were captured using the IncuCyte FLR in build software, darker region represents computed wound barrier. **Figure Overleaf.**

Figure 4.5.2– TamR cells have a migratory phenotype

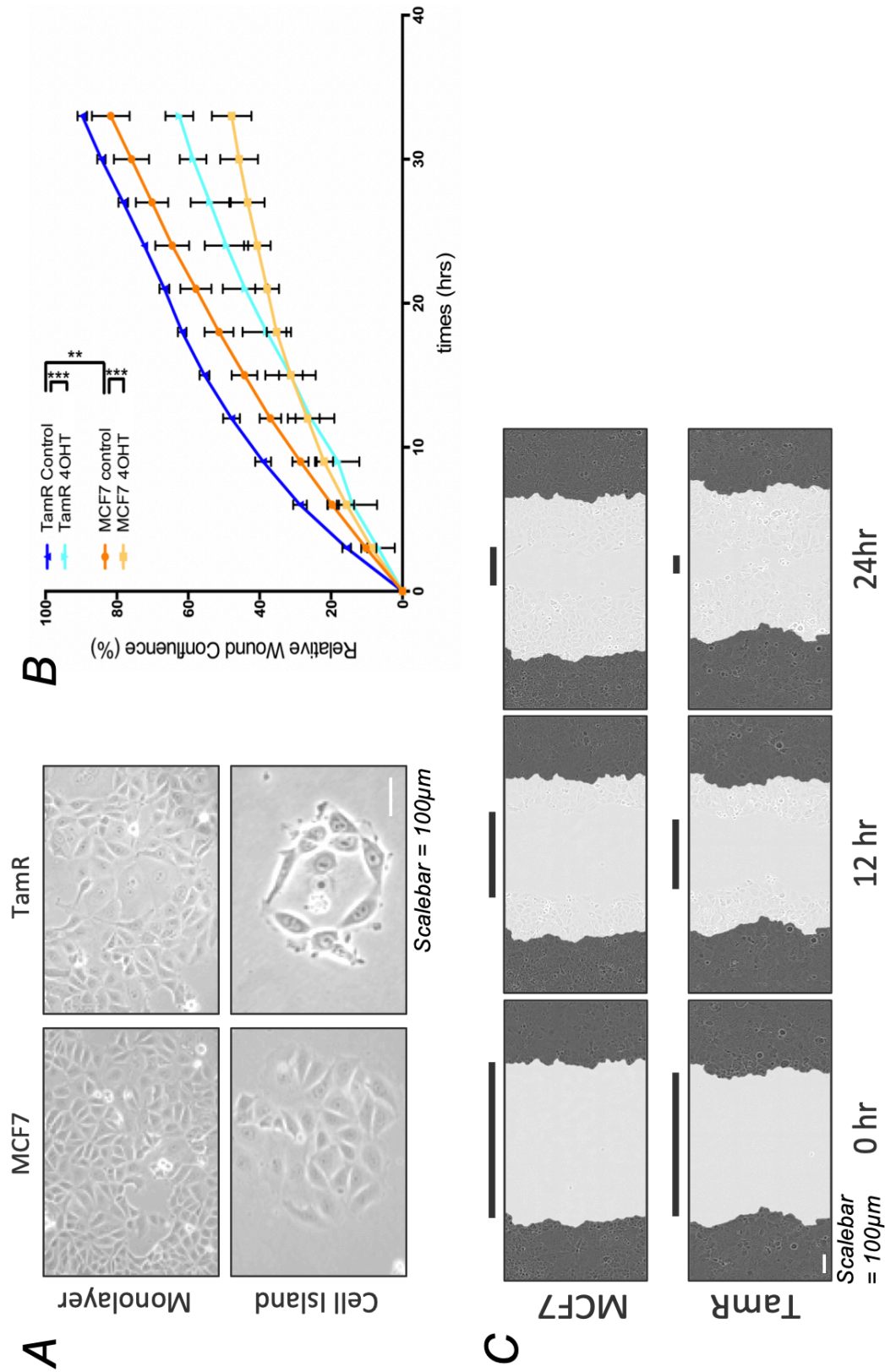
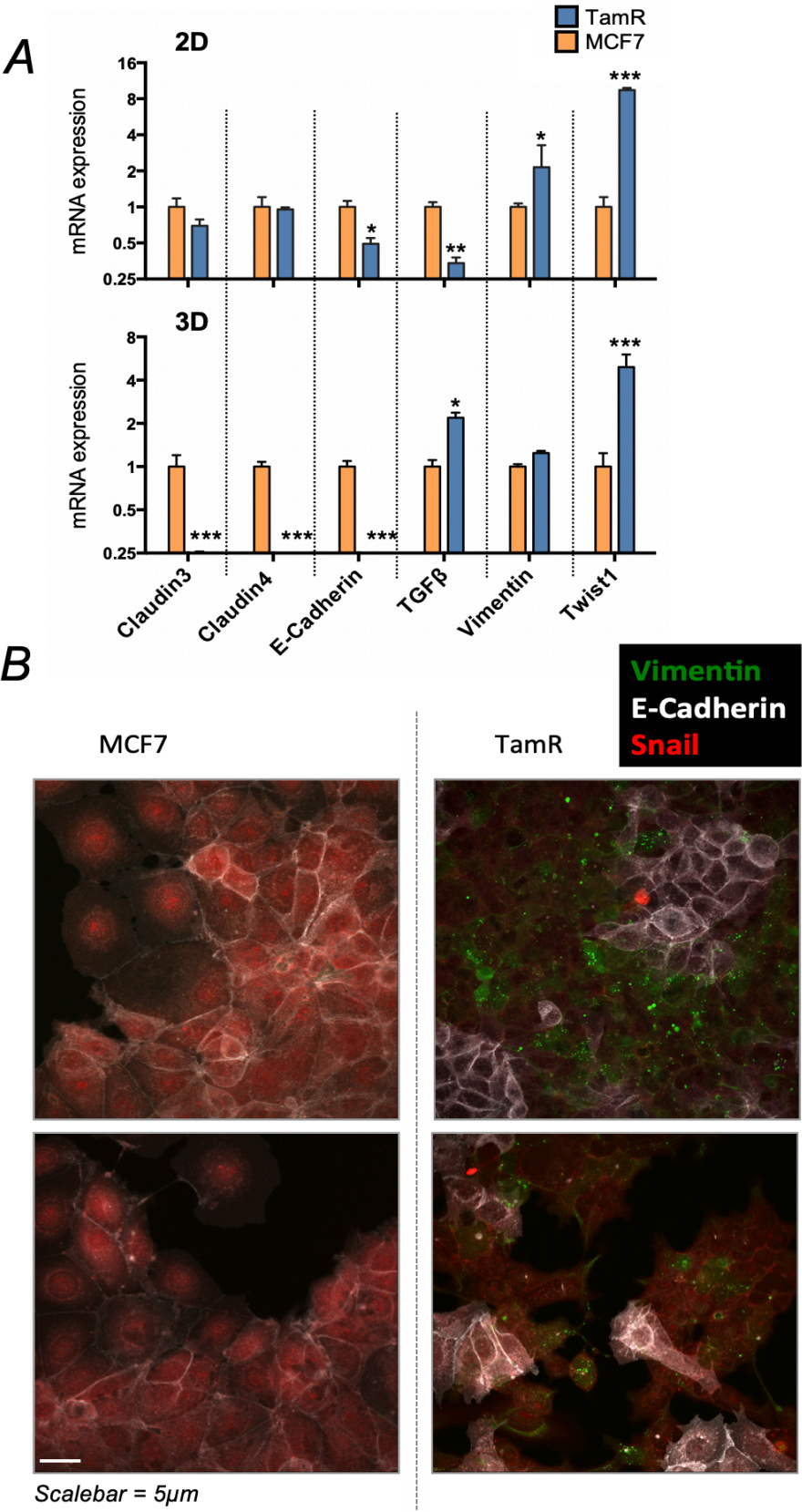


Figure 4.5.3 – TamR cells have some features of EMT

A) qPCR profiling of EMT markers in each cell line and in 2D (monolayer) and 3D (mammosphere) culture. Relative expression is computed by comparative C_T method ($2^{-\Delta\Delta C_T}$) (333) to the geometric mean of B2M and GAPDH, and MCF7 expression. Significance was calculated by one-way ANOVA with Bonferroni to correct for multiple comparisons (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$). **B)** Immunofluorescence of key EMT regulatory and indicator proteins: Vimentin (pseudocolored green), E-Cadherin (pseudocolored white) and SNAI1 (pseudocolored red) in each cell line.

Figure Overleaf.

Figure 4.5.3 – TamR cells have some features of EMT



4.6 *in vivo* characteristics of OCT4-Vex-Poly MCF7s

We have previously found that MCF7 OCT4-Vex-Poly is markedly slower growing *in vitro* than its parental line (*Figure 4.4.2*). We have also found evidence of ER α dependent growth in this line, in terms of Ki67 and ER α staining overlap by IHC of mammospheres (*Figure 4.5.1*). We sought to investigate the tumour forming ability, and oestrogen dependence, of OCT4-Vex-Poly compared to parental lines. *Figure 4.6.1 A* shows *in vivo* growth curves of subcutaneously injected MCF7 and OCT4-Vex-Poly cell lines in NSG mice with slow release human oestrogen pellets to aid in tumour growth. OCT4-Vex-Poly tends to form palpable tumours in a high oestrogen environment at a slightly slower rate than parental MCF7s (*Figure 4.6.1 A*). *Figure 4.7.1 B* shows tumours harvested from the flanks of NSG mice 80 days following implantation, reasonable concordance between palpable tumour size and actual volume can be confirmed visually, as can the slower growth rate of MCF7 OCT4-Vex-Poly tumours relative to parental. Next, we repeated this experiment in non-oestrogen supplemented NSGs. Interestingly, in this low oestrogen environment, OCT4-Vex-Poly cell lines form palpable tumours at a markedly faster rate than MCF7 parental (*Figure 4.6.2 A*). At 110 days the experiment was ended, and tumours harvested, *Figure 4.6.2 B* shows freshly harvested tumours from each arm.

IHC of OCT4-Vex-Poly tumours grown in *Figure 4.6.1* reveal a markedly more fibrous structure in OCT4-Vex-Poly versus parental MCF7 (*Figure 4.6.3*). GFP is clearly marked, and is heterogeneous in localisation, this could reflect the polyclonal nature of our derived cell line, a preferential localisation of *POU5F1* expressing cells in the tumour, or simply the fibrous nature of the tumours. Interestingly, OCT4-Vex-Poly tumours maintain ER α expression comparable to MCF7 parental tumours and there

appears to be a co-localisation with Ki67, suggesting the maintenance of oestrogen or ER α dependent growth.

Next, we sought to investigate the presence of BC-TICs in OCT4-Vex-Poly and parental MCF7s, with and without exposure to 4-OHT. Thus, we turned to the *in vivo* version of the LDA assay, injecting increasingly dilute numbers of pre-treated cells into the flanks of oestrogen supplemented NSG mice. Tabulated and graphical results are shown in (Figure 4.6.4). Statistical analysis by ELDA revealed estimates of 1/87.70 BC-TICs in a pool of MCF7s, raising to 1/14.33 when pre-treated with 4-OHT ($p=0.0103$). In OCT4-Vex-Poly BC-TIC frequency was 1/35.14, rising to 1/4.81 with the addition of 4-OHT ($p=0.0016$). Thus, 4-OHT is found to enrich for BC-TICs in both parental and OCT4-Vex-Poly cell lines.

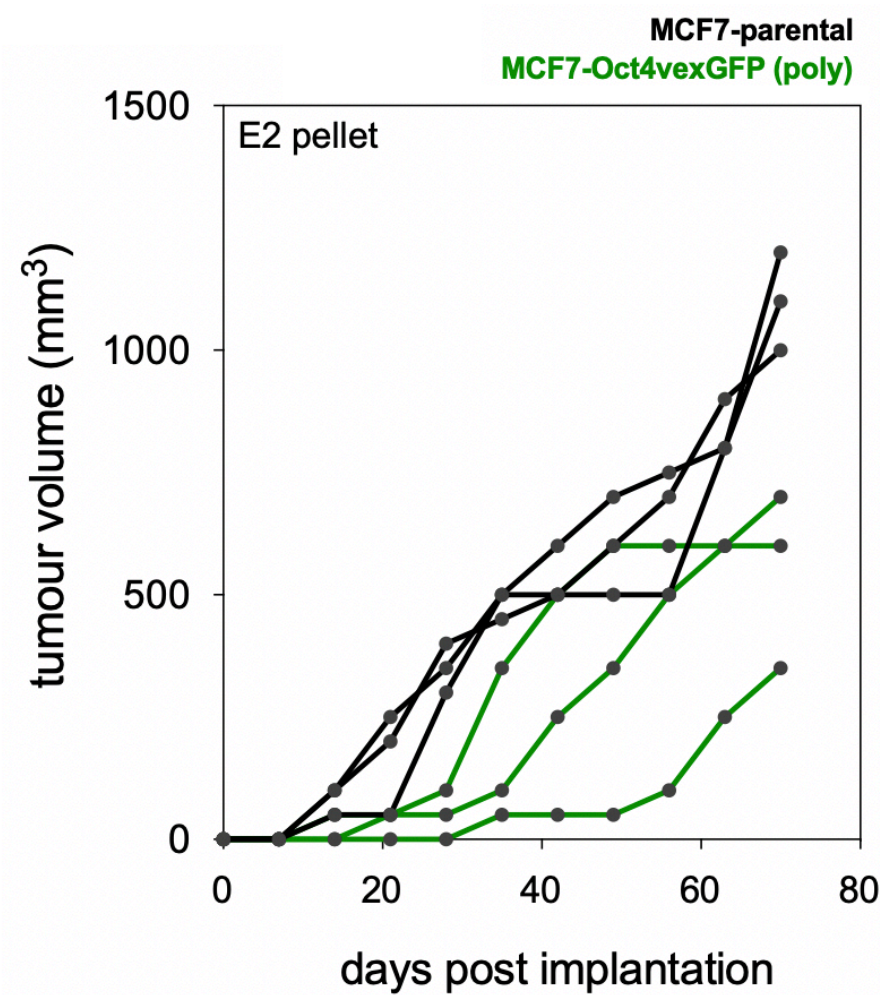
These results are concordant with our *in vitro* surrogate assays. *POU5F1* over-expression significantly enhances the selection of BC-TICs by 4-OHT. It may be that OCT4 activity is a feature of the phenotypic clones responsible for the development of tamoxifen resistance defined in Chapter 3.

Figure 4.6.1 – Characterization of tumour growth in OCT4-Vex-Poly MCF7s

A) *In vivo* tumour growth comparison of OCT4-Vex-Poly versus MCF7 parental lines implanted into the flank of oestrogen-pellet-supplemented NSG mice. Tumour size was estimated periodically by palpation. OCT4-Vex-Poly derived tumours tended to be slower growing in E2 supplemented NSGs relative to parental MCF7s. **B)** 70-day tumours were harvested and measured, reasonable concordance with final palpitation estimated size is seen by visual inspection. **Figure Overleaf.**

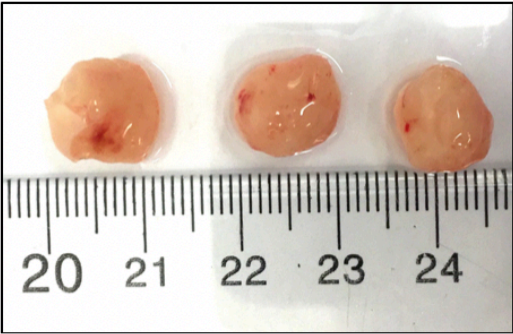
Figure 4.6.1 – Characterization of tumour growth in OCT4-Vex-Poly MCF7s

A



B

MCF7 Parental



MCF7 Oct4-Vex-Poly

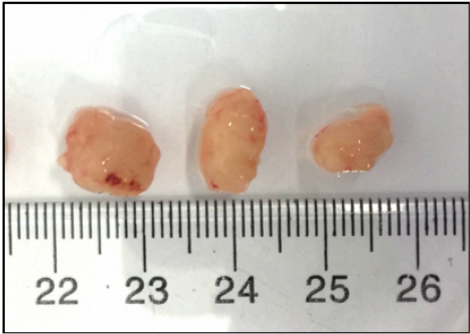


Figure 4.6.2 – E2 dependence in OCT4-Vex-Poly MCF7 tumours

A) OCT4-Vex-Poly and MCF7 cells were injected into the flank of NSG immunocompromised mice, without oestrogen pellets, to assay tumour growth in a low oestrogen environment. OCT4-Vex-Poly appeared to initiate tumours more effectively than MCF7 in the low oestrogen environment, with markedly smaller tumours at 100 days in parental versus POU5F1 overexpressing. **B)** Tumour volume estimates by palpitation are validated visually by 100-day harvested tumours. **Figure Overleaf.**

Figure 4.6.2 – E2 dependence in OCT4-Vex-Poly MCF7 tumours.

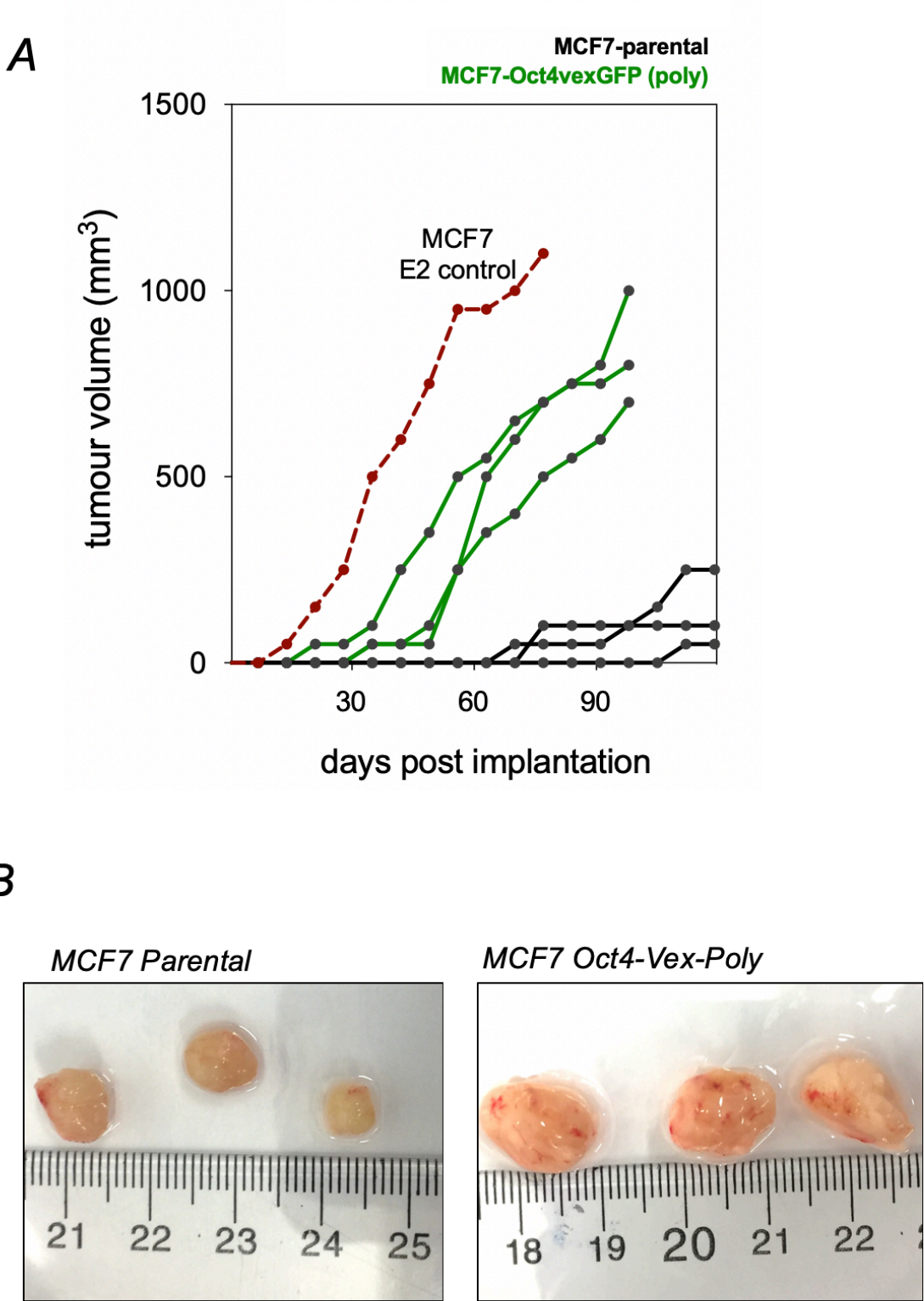


Figure 4.6.3 – IHC of OCT4-Vex-Poly MCF7 tumours

OCT4-Vex-Poly and MCF7 parental-derived tumours from Figure 4.5.2 were stained by IHC for GFP, ER α & Ki67 and morphology investigated by H&E. OCT4-Vex-Poly exhibited a fibrous tumour morphology by H&E, with reduced cellularity relative to MCF7 parental. GFP is clearly stained, confirming OCT4-Vex-Poly integration and continued transgene expression, though expression appears heterogeneous in localisation. OCT4-Vex-Poly tumours maintain ER α expression comparable to MCF7 parental tumours and there appears to be a co-localisation with Ki67, suggesting that these cells proliferate in response to ER α -pathway activation or at least ER α presence.

Figure Overleaf.

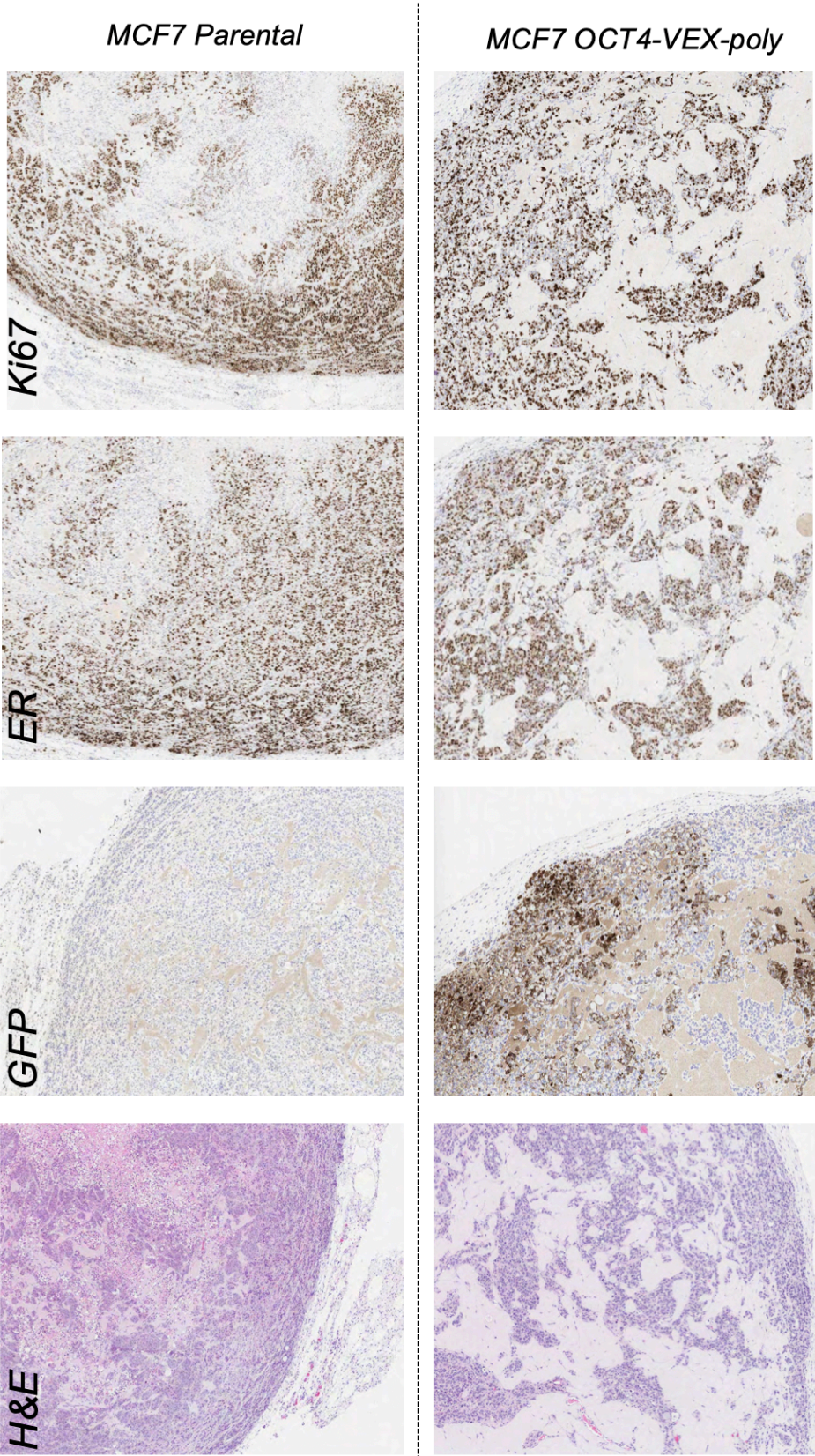
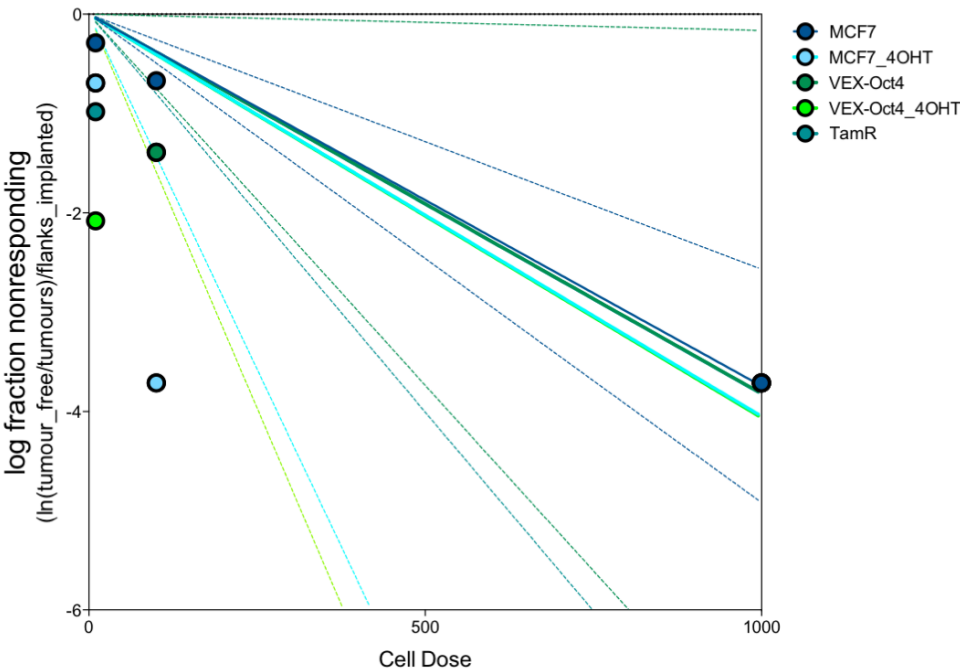


Figure 4.6.3 – IHC of OCT4-Vex-Poly MCF7 tumours

Figure 4.6.4 – in vivo LDA of OCT4-Vex-Poly MCF7s

*Shows results of an in vivo LDA. Following pre-treatment with 4-OHT, where appropriate (24 hrs at 200nM), cells were harvested, serially diluted and resuspended in 50% Matrigel. 50µl dilutions containing 1,000, 100 or 10 cells as appropriate were injected into the 4th mammary fat pads of NSG mice. NSGs were also injected with 0.72mg/90day slow release E2 pellets. Tumour injection sites were palpated for signs of tumour growth twice weekly for up to 6 months. Statistical analysis by ELDA revealed estimates of 1/87.70 BC-TICs in a pool of MCF7s, raising to 1/14.33 when pre-treated with 4-OHT ($p=0.0103$). In OCT4-Vex-Poly BC-TIC frequency was 1/35.14, rising to 1/4.81 with the addition of 4-OHT ($p=0.0016$). These results again highlight the co-operative nature of OCT4 and 4-OHT in enriching for BC-TIC-like cells. **Figure Overleaf.***

Figure 4.6.4 – *in vivo* LDA of OCT4-Vex-Poly MCF7



Cell Dose	MCF7	MCF7_4OHT	VEX-Oct4	VEX-Oct4_4OHT	TamR
1000	4 in 4	4 in 4	4 in 4	4 in 4	4 in 4
100	2 in 4	4 in 4	3 in 4	4 in 4	3 in 4
10	2 in 8	4 in 8	4 in 8	3 in 4	5 in 8

Group	Lower	Estimate	Upper
MCF7	257.8	87.79	29.89
MCF7_4OHT	38	14.33	5.41
TamR	78	28.79	10.63
VEX-Oct4	94	35.14	13.13
VEX-Oct4_4OHT	11.6	4.81	1.99

Group 1	Group 2	Chisq	DF	Pr(>Chisq)
MCF7	MCF7_4OHT	6.58	1	1.030E-02
MCF7	TamR	3.1	1	7.830E-02
MCF7	VEX-Oct4	1.97	1	1.600E-01
MCF7	VEX-Oct4_4OHT	18.7	1	1.550E-05
MCF7_4OHT	TamR	1.15	1	2.840E-01
MCF7_4OHT	VEX-Oct4	1.82	1	1.770E-01
MCF7_4OHT	VEX-Oct4_4OHT	2.75	1	9.730E-02
TamR	VEX-Oct4	0.116	1	7.330E-01
TamR	VEX-Oct4_4OHT	8.41	1	3.730E-03
VEX-Oct4	VEX-Oct4_4OHT	9.96	1	1.600E-03

4.7 Tamoxifen Resistant PDX Models

From our established biobank of breast cancer PDX models (198), we next selected 14 ER α + models for further study and confirmed their resistance to 4-OHT. Each model was dissociated into PDTs for short term *ex vivo* culture as previously described, (198). PDTs were then treated with an escalating dose of 4-OHT and IC50s were calculated. *Figure 4.7.1 A* shows combined dose response curves for each model, with IC50 values (in Molar concentration) in the figure legend. We have previously calculated area under the curve (AUC) from dose response curves of multiple PDX models for multiple drugs and combination therapies by high throughput screens (HTS) (198). Eight models had data available for 4-OHT treatment. *Figure 4.7.1 B* shows good correlation between IC50s calculated in these experiments and previously found AUCs ($R^2 = 0.6533$; $p = 0.0047$), hence we interpolated AUCs from models not tested in HTS. Results were as follows: IC06-x5 0.361; IC06_TAMR-x4 0.206; VHIO131-x4 0.311; HCI006-x4 0.343; STG335-x5 0.294. We took a cut-off of AUC 0.2 to designate models as resistant to 4-OHT, based on analysis of our HTS (198). We included any model with IC50 of over 1×10^{-6} M, representing the five most right shifted dose response curves in *Figure 4.7.1*. Thus, eight PDX models were deemed sensitive to tamoxifen (STG195, STG335, HCI002, HCI005, HCI006, VHIO098, VHIO131 & VHIO244), and six deemed resistant (STG143, IC06, IC06_TAMR, HCI011, STG201, & IC07).

As discussed, resistance to tamoxifen has been linked to expression of several master transcription factors typically associated with guiding embryonic stem cell fate and differentiation. We profiled each selected PDX model for embryonic transcription factors OCT4 (*POU5F1*), *SOX2*, *NANOG* & *KLF4* alongside putative BC-TIC markers *CD44* and *PROCR* (167,402). *Figure 4.7.2* shows the results of this analysis. Several markers appear to be enriched in resistant models. This is clearest in *POU5F1* and *SOX2*, though in *Figure 4.7.3* we see several differences between the

mean expression for each marker in each model. Significant differences are seen between sensitive and resistant groups (by one-way ANOVA with Bonferroni correction) for *POU5F1* ($p=0.0065$) and *SOX2* ($p=0.0412$) expression.

Aldehyde dehydrogenase (ALDH) activity is a key marker of mammary stem cells and BC-TICs (161,166). High ALDH expression has been reported for normal and cancer precursor cells of various lineages, including hematopoietic, mammary, endothelial, mesenchymal, and neural (428). We next used an ALDH activity assay (ALDEFLUOR, see *Methods, section 2.3.6*) to assay PDX models, treated with 200nM 4-OHT for 24hrs by flow cytometry. *Figure 4.7.4 A* shows a representative flow cytometry plots from STG335x3, with greatly increased numbers of ALDH⁺ cells seen after acute treatment with 4-OHT. Overall, we found three out of six resistant models had significantly reduced ALDH⁺ cells following treatment. Conversely, six out of eight sensitive models were significantly enriched for ALDH⁺ cells following treatment (*Figure 4.7.4 B*).

Next, we used the ALDH activity assay to derive separate, and potentially BC-TIC enriched, populations from four PDX models. *Figure 4.7.5 A* shows a flowcytometry gating strategy to derive ALDH⁺ and ALDH⁻ populations from STG335-x4. Flow sorted cells were separately plated and grown for up to seven days *ex vivo* as PDTs, and 4-OHT dose response analyses were performed. We find that resistant PDX models (ICO6 & STG143) have little difference between 4-OHT sensitivity in ALDH⁺ and ALDH⁻ populations, but that 4-OHT sensitive models (STG195 & STG335) show significant disparity in the drug resistance of ALDH cell-sorted populations. In both cases, the ALDH⁺ population, is significantly more resistant to 4-OHT treatment than either the bulk population or the ALDH⁻ cells (by one-way ANOVA with Bonferroni correction). These data are consistent with our cell line studies; populations enriched for ALDH⁺ cells or M-ICs, have a higher 4-OHT tolerance than their non-enriched counterparts.

Figure 4.7.1 – 4-OHT resistance in PDX models

A) Dose response data for 4-OHT in 14 ER α + breast cancer PDX models, cultured short term ex vivo as PDTs. IC₅₀ values are computed and shown in the legend. There is a 10-fold spread in IC₅₀ values across the models, STG195 is the most sensitive to 4-OHT with an IC₅₀ of 0.19 μ M, IC07 is the most resistant with an IC₅₀ of 1.8 μ M. **B)** Area under the curve (AUC) analysis from our high throughput (HTS) PDX drug screens (198) is compared to IC₅₀ from (A). A good correlation is seen, $R^2=0.6533$ $p=0.0047$, and missing AUC values at interpolated. Results are as follows: IC06 x5 0.361; IC06_TAMR x4 0.206; VHIO131 x4 0.311; HCI006 x4 0.343; STG335 x5 0.294. **Figure Overleaf.**

Figure 4.7.1 – 4-OHT resistance in PDX models

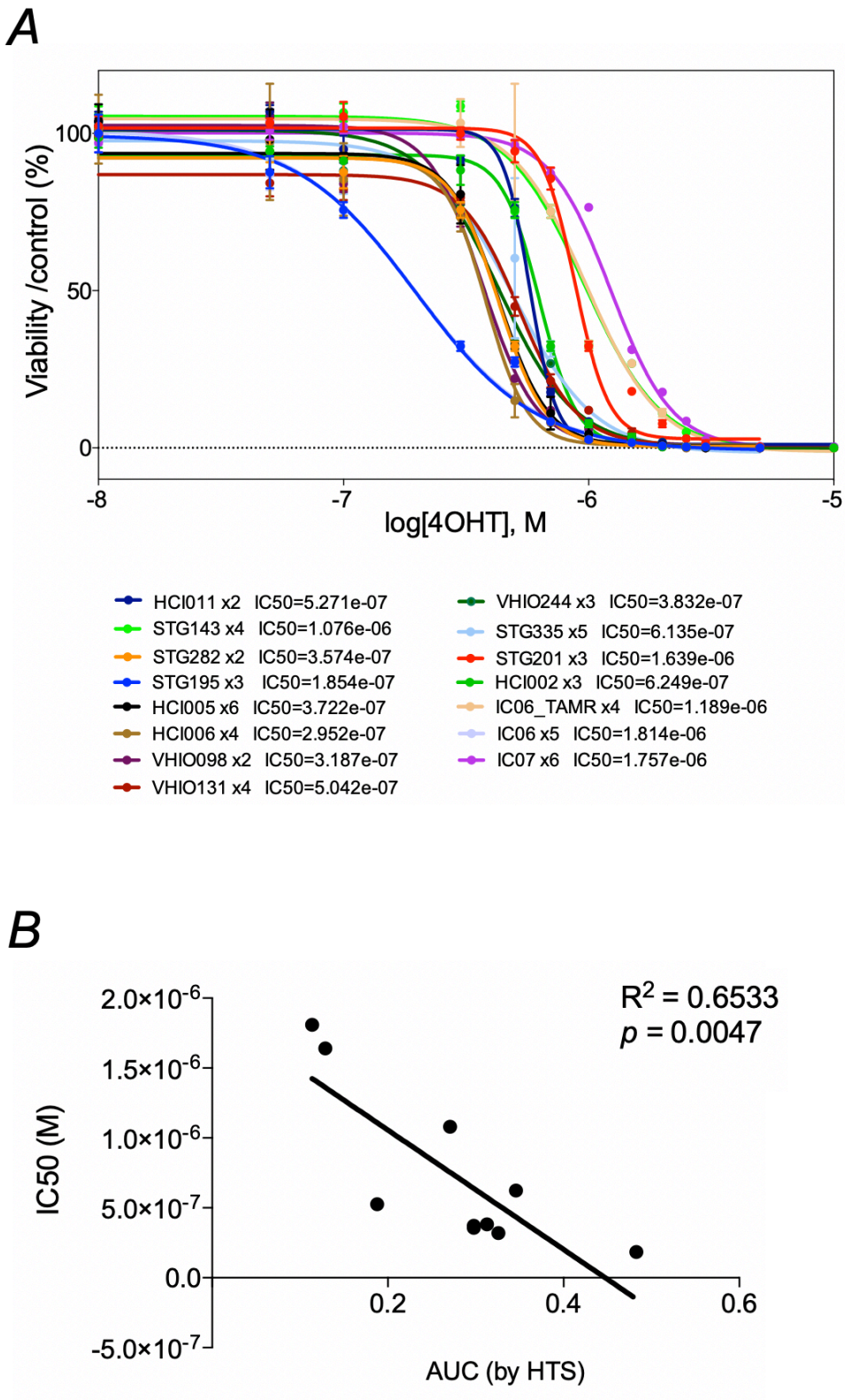


Figure 4.7.2 – qPCR for stem cell markers

*Stem cell transcription factors NANOG, POU5F1, SOX2 and KLF4 are profiled across tamoxifen resistant and sensitive models by qPCR. Putative BC-TIC surface markers PROCR and CD44 are also profiled. Results are normalized by comparative C_T method ($2^{-\Delta\Delta C_T}$) (333) to the geometric mean of B2M and GAPDH, and STG195 expression (see Methods, section 2.4.1). **Figure Overleaf.***

Figure 4.7.2 – qPCR for stem cell markers

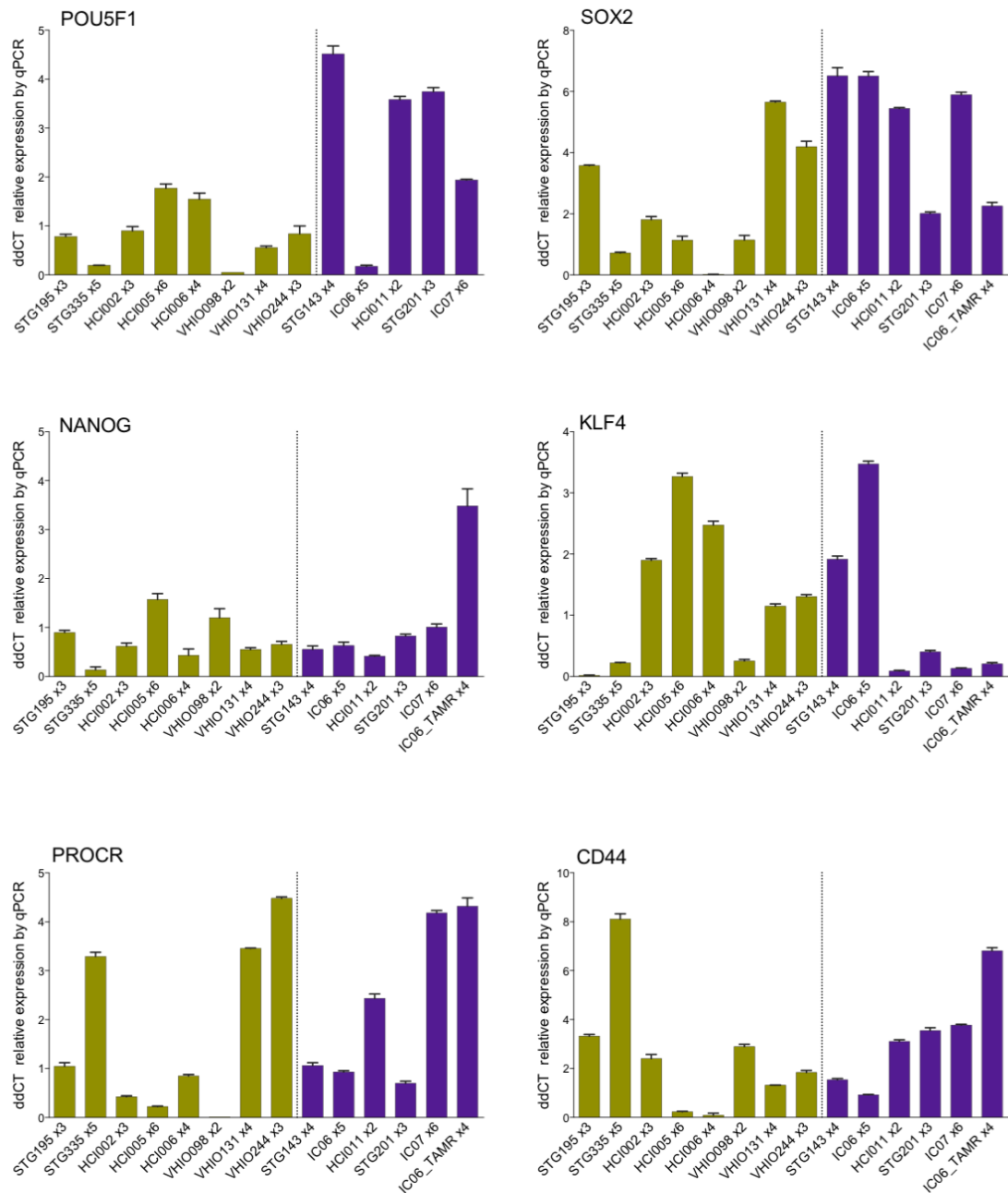


Figure 4.7.3 – Combined analysis of qPCR for stem cell markers

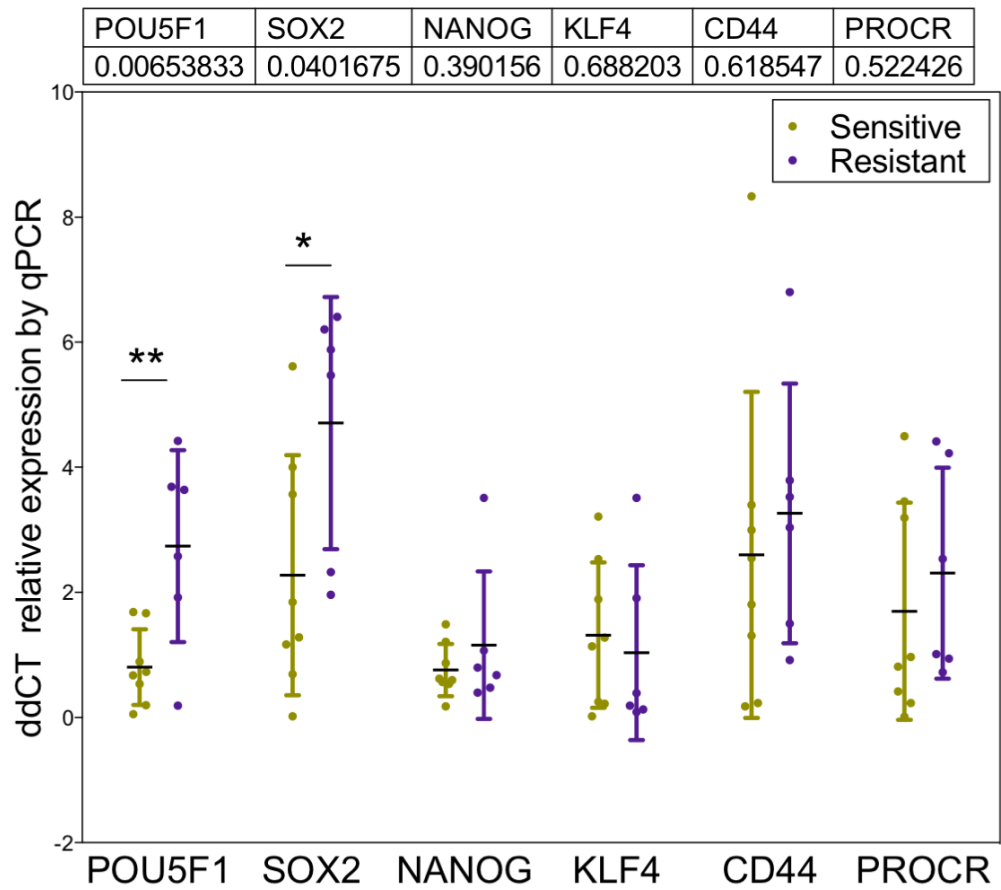


Figure 4.7.3 – Combined analysis of qPCR for stem cell markers

Mean expression values relative to MCF7 from Figure 4.7.2 were combined for resistant and sensitive groups. POU5F1 & SOX2 were significantly upregulated in resistant versus sensitive models. Statistical significance was computed by one-way ANOVA with Bonferroni correction and is presented versus MCF7 (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$).

Figure 4.7.4 – Tamoxifen ALDH+ selection

*PDTCs were assayed for ALDH activity by flow cytometry in the ALDEFLUOR assay (see Methods, section 2.3.6). **A)** shows representative flow cytometry plots showing ALDH activity increasing in STG335 after 24hrs of 200nM 4-OHT. **B)** Shows this assay repeated across PDTC models. Overall, we found 3 out of 6 resistant models had significantly reduced ALDH+ cells following treatment. Conversely, 6 out of 8 sensitive models were significantly enriched for ALDH+ cells following treatment. Statistics computed by one-way ANOVA with Bonferroni correction for multiple comparisons (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$). **Figure Overleaf.***

Figure 4.7.4 – Tamoxifen ALDH+ selection

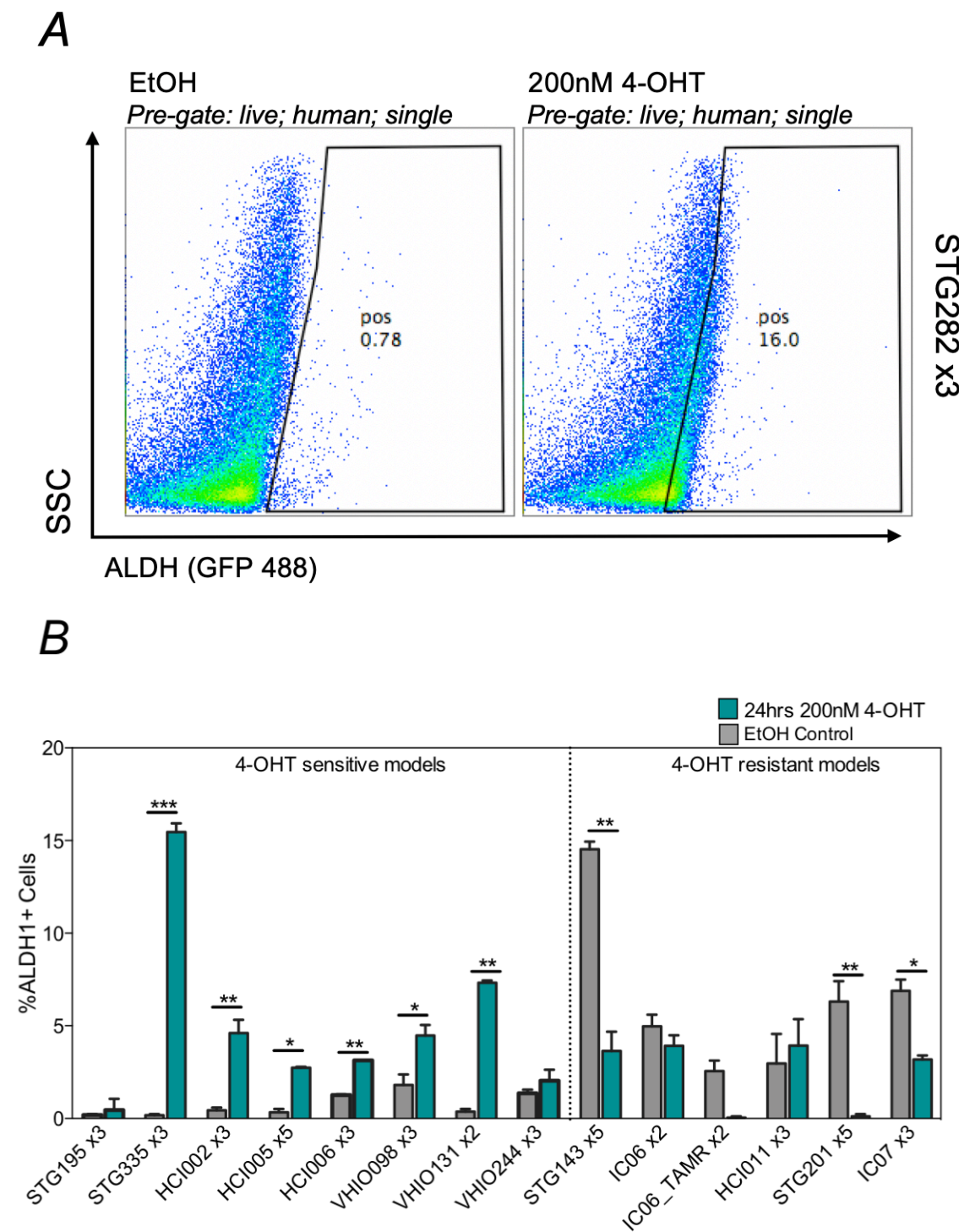
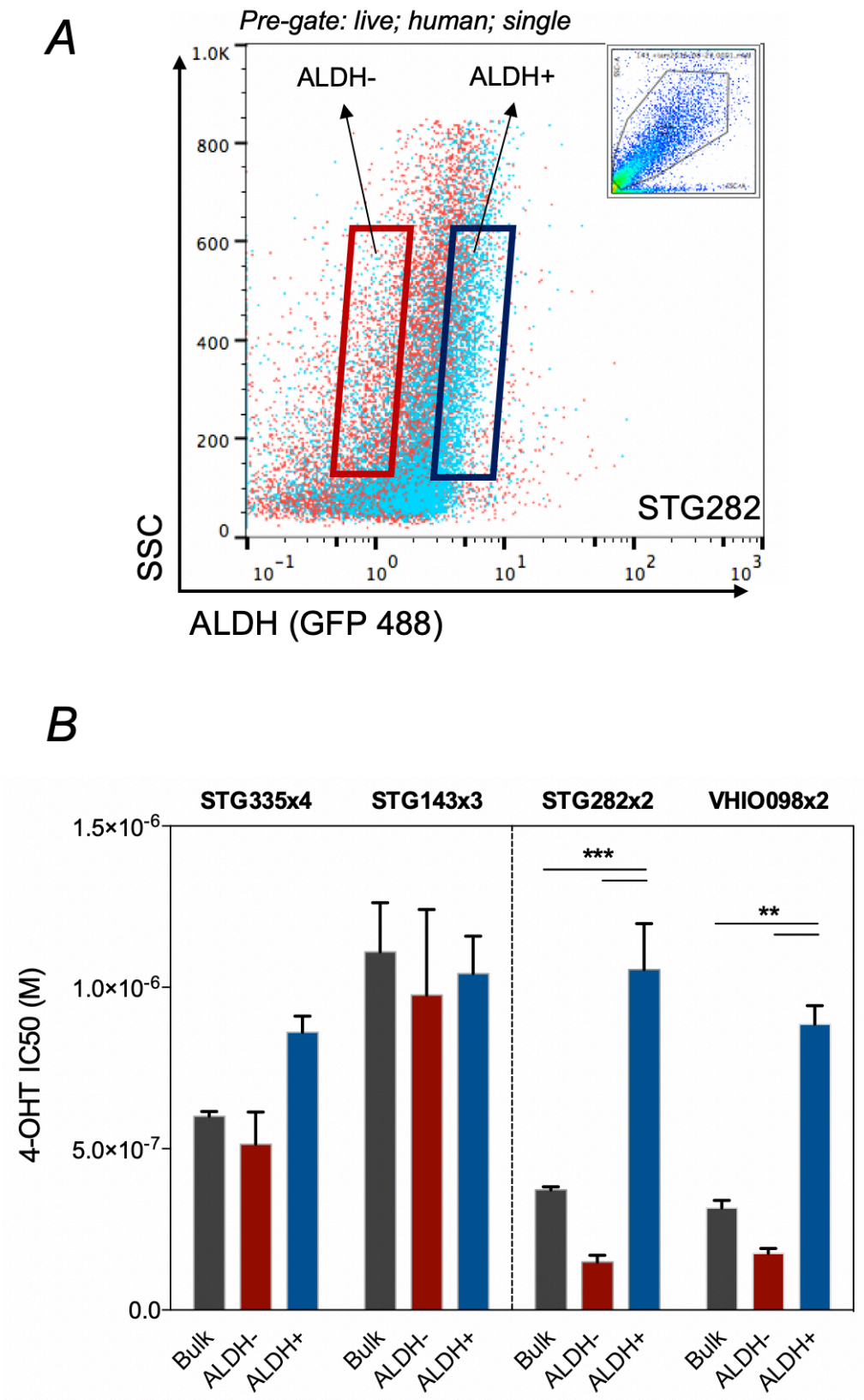


Figure 4.7.5 – ALDH⁺ sorted PDTCs are resistant to 4-OHT

A) PDTCs isolated from four PDX models were assayed for ALDH activity. Gating strategy is shown for selection of high ALDH activity (ALDH⁺) and low ALDH activity (ALDH⁻) populations. ALDH⁺ and ALDH⁻ cells were flow sorted and individually culture. **B)** ALDH⁺, ALDH⁻ and bulk PDX cultures were assayed for 4-OHT sensitivity by dose response assay. In both 4-OHT sensitive models (STG195 & STG335), ALDH⁺ populations were significantly more resistant to 4-OHT than either bulk or ALDH⁻ cells. In resistant models (IC06 & STG143), there were no significant differences in sensitivity to 4-OHT between populations. Significance was computed by one-way ANOVA with Bonferroni correction (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$). **Figure Overleaf.**

Figure 4.7.5 – ALDH+ sorted PDTCs are resistant to 4-OHT



4.8 BC-TIC Markers in the METABRIC cohort

Having shown significant enrichment of markers for BC-TIC activity in our tamoxifen resistant models, we sought to understand the potential clinical impact by interrogating the METABRIC dataset of >2000 breast cancer patients (7,73). Several differences exist in marker expression across the 11 integrative clusters (ICs) of breast cancer, *Figure 4.8.1*. For example, *SOX2* appeared to be enriched in IC1 (ER α +) & IC5 (HER2+ & ER α + / HER2+). Interestingly, the opposite pattern is seen with *CD44*, which had the lowest expression in IC1 & IC5. *KLF4* and *PROCR* were highly variably expressed between the clusters, with *KLF4* appearing to show some enrichment in IC3 and *PROCR* in IC4 ER α -. *POU5F1* and *NANOG* are more stable across the clusters, with no obvious bias. Surprisingly, we find none of the BC-TIC markers or transcription factors were enriched in IC10, a cluster highly enriched for TNBC, typically thought to be dedifferentiated and BC-TIC rich (173,429–431). *NANOG* in particular has recently been linked to TNBC, with one study reporting that the connexin Cx26 can promote BC-TIC self-renewal by forming a signalling complex between *NANOG* and focal adhesion kinase (FAK) (429).

Next, we isolated ER α + patients who had gone through hormonal therapy from the METABRIC dataset. Surprisingly, we find *CD44* depleted in ER α + patients who have gone through hormonal therapy, but not in the ER α + subgroup as a whole (*Figure 4.8.2 A*). Both *PROCR* and *KLF4* were significantly enriched in both ER α + and hormone-treated patient groups ($p < 0.05$; one-way ANOVA with Bonferroni correction). However, taking overexpression as a z-score threshold of 2.0 (from Illumina Human v3 microarray), we find relatively little difference in overall survival for patients enriched for any of our BC-TIC markers versus the ER α + population as a whole (*Figure 4.8.2 B*). Though differences do exist in median overall survival, for example with *POU5F1* overexpression translating to 151 months median survival ($n=82$) compared to 162 months median survival for ER α + as a

whole ($n=1515$). Overall, the survival curves were not significantly different by the Logrank test for trend, with $p=0.61$.

To further investigate the role of OCT4 in tamoxifen resistance, we compared *POU5F1* expression in ER α + patients from the METABRIC cohort who had not responded to their hormone therapy and ultimately died of their cancer, versus those treated with hormone therapy and still alive. Interestingly, we found significantly ($p=0.0052$; two-tailed t-test) increased *POU5F1* expression in those who had died ($n=393$) of ER α + hormone-treated breast cancer compared to those still alive ($n=495$) (*Figure 4.8.3 A*). Further, patients in the ER α + hormone-treated group with high *POU5F1* expression (in the top100) had worse overall survival than those with low *POU5F1* expression (bottom100) ($p = 0.045$; Logrank (Mantel-Cox) test) (*Figure 4.8.3 B*). These analyses highlight the potential role of BC-TICs expressing embryonic transcription factors such as Oct4 in the development of tamoxifen resistance and in poor-outcome ER α + /hormone treated breast cancer.

Figure 4.8.1 – BC-TIC phenotype in the METABRIC dataset

*Log Intensity Expression values from Illumina Human v3 microarray of BC-TIC marker expression across the 2000 breast cancer patients of the METABRIC dataset are presented for SOX2, POU5F1, KLF4, NANOG, PROCR & CD44. Several trends toward Integrative Cluster (IC) enrichment were observed. For example, SOX2 appears to be enriched in IC1 & IC5, CD44 is lowest expression in IC1 & IC5, KLF4 & PROCR are highly variable between the clusters. **Figure Overleaf.***

Figure 4.8.1 – BC-TIC phenotype in the METABRIC dataset

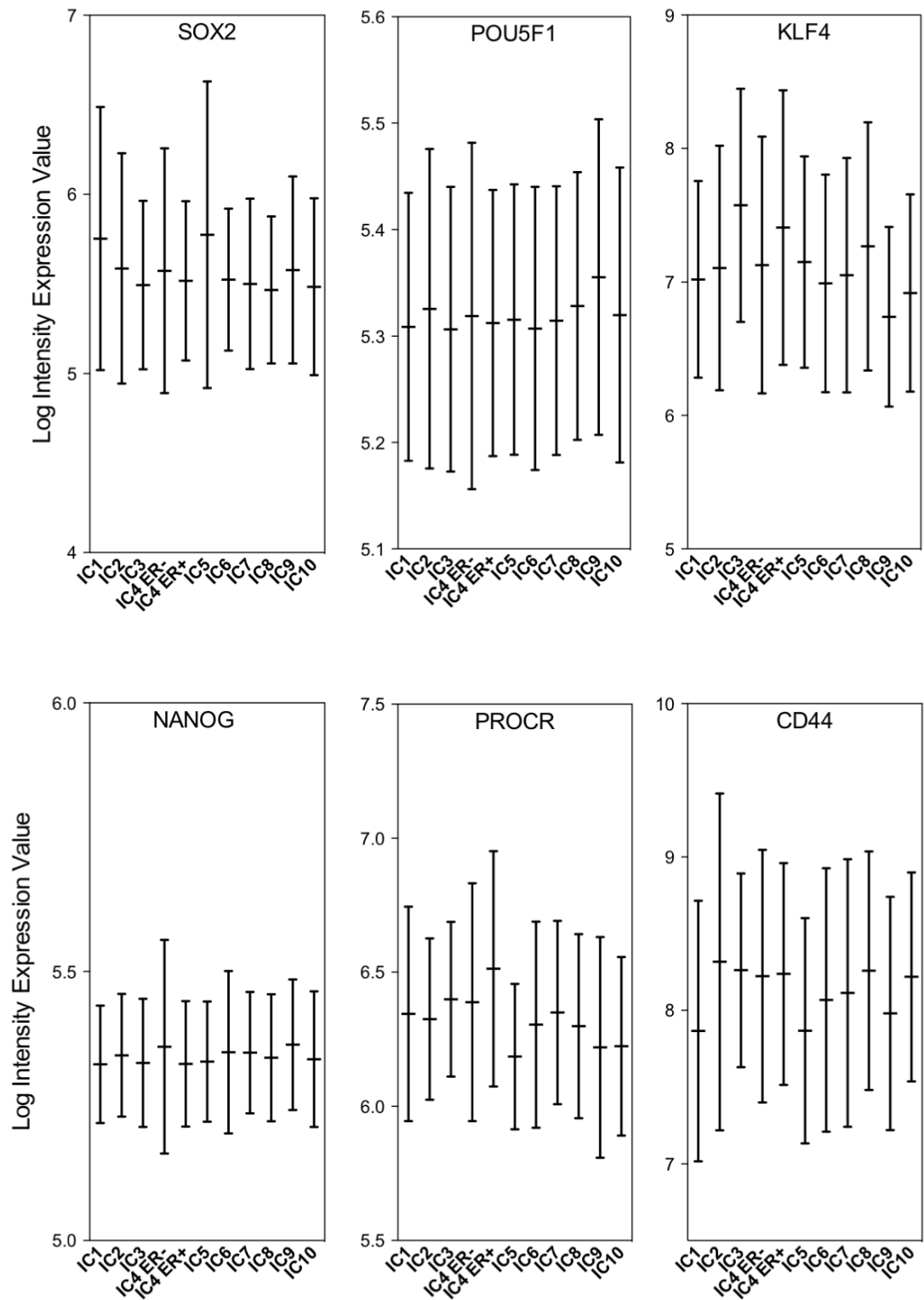
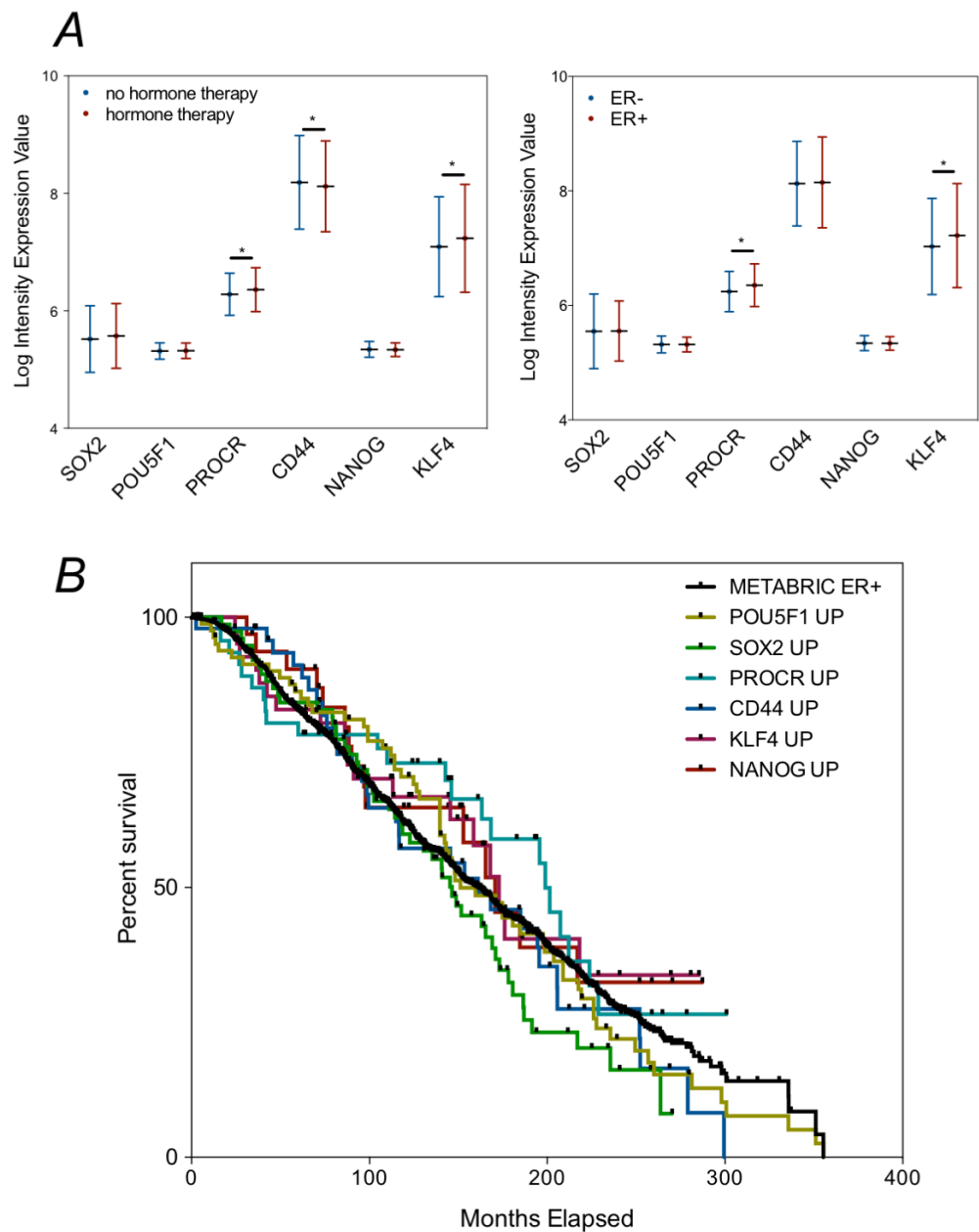


Figure 4.8.2 – METABRIC survival

A) *SOX2, POU5F1, KLF4, NANOG, PROCR & CD44 expression is compared in METABRIC patients with hormone/no hormone therapy and those who are ER α + and ER α -. CD44 expression was significantly (though not markedly) reduced in ER α + patients who have gone through hormonal therapy, but not in the ER α + subgroup as a whole, PROCR and KLF4 were significantly enriched in both ER α + and hormone-treated patient groups ($p < 0.05$; one-way ANOVA with Bonferroni correction). **B)** Taking overexpression as a z-score threshold of two, we find relatively little difference in overall survival for patients enriched for any of our BC-TIC markers versus the ER α + population as a whole by Kaplan-Myer analysis. **Figure Overleaf.***

Figure 4.8.2 – METABRIC survival



Median Overall Survival (Months)

METABRIC ER+	POU5F1 UP	SOX2 UP	PROCR UP	CD44 UP	KLF4 UP	NANOG UP
161.667	151.2	146.4	199.033	161.133	172.8	170.667

Number of Patients in Group

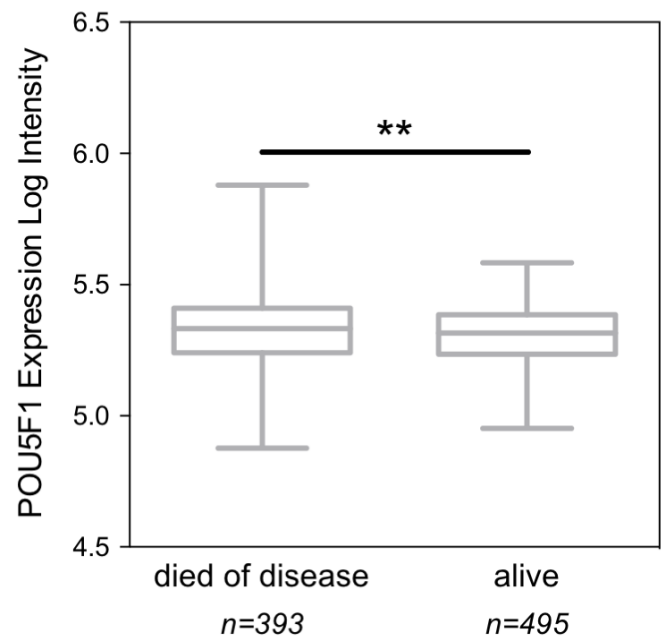
METABRIC ER+	POU5F1 UP	SOX2 UP	PROCR UP	CD44 UP	KLF4 UP	NANOG UP
1515	82	77	47	48	44	32

Figure 4.8.3 – POU5F1 in METABRIC

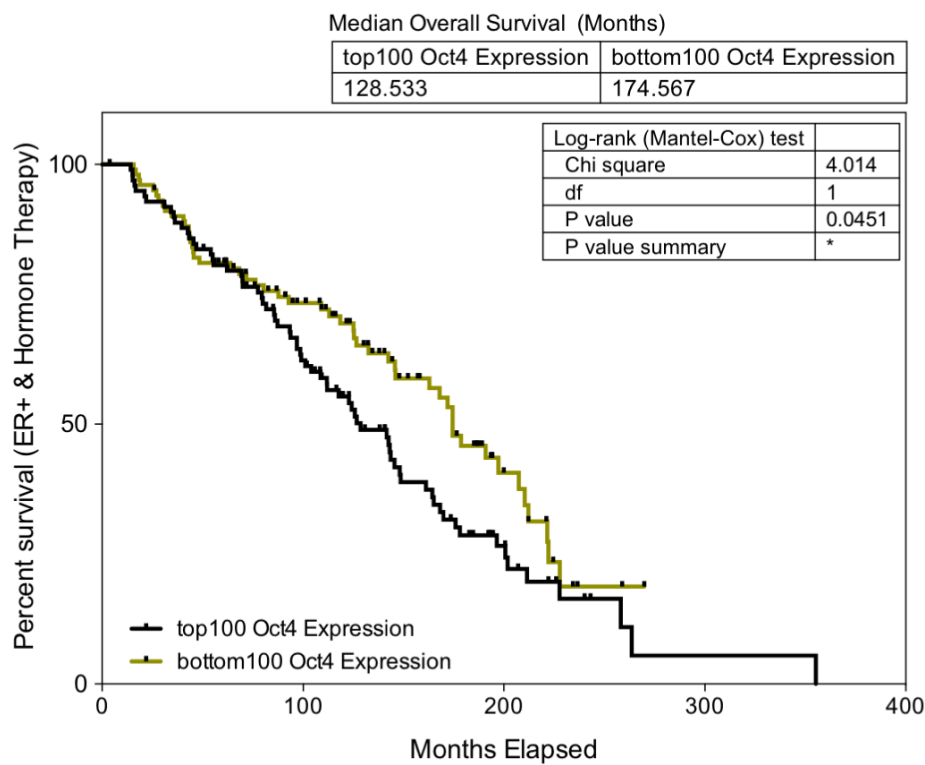
A) POU5F1 is moderately higher expressed in ER α + patients who have died from their disease versus those who are alive ($p=0.0052$; two-tailed t -test). **B)** POU5F1 has prognostic value in ER α + patients undergoing hormonal therapy, high POU5F1 expression is significantly ($p=0.045$) associated with worse overall survival in this patient group ($p = 0.045$; Logrank (Mantel-Cox) test). **Figure Overleaf.**

Figure 4.8.3 – POU5F1 in METABRIC

A



B



4.9 General Discussion

In this chapter, we have experimentally investigated the link between tamoxifen resistance and a CSC-like phenotype, described *in vitro* by M-IC assays and *in vivo* by BC-TIC assays. We uncover evidence that treatment with 4-OHT enriches for both M-ICs and BC-TICs in initially tamoxifen-sensitive cell line and PDX models. This is verified in functional BC-TIC assays *in vivo*, as well as in characterising expression of canonical embryonic stem cell transcription factors. M-IC enrichment is likely through selective cell death, but there is a suggestion that 4-OHT-ER α is perhaps able to induce *POU5F1* expression directly. We find that OCT4 is essential for the M-IC enrichment process, but not for normal M-IC maintenance, and find a plausible mechanism through which ER α or 4-OHT-ER α may directly induce the expression of *POU5F1*. Much more work is required to uncover and validate a comprehensive molecular mechanism to explain all of our results, however, the results herein are a valuable step in beginning to explain apparent contradictions in the development of tamoxifen resistance described in *Chapter 3*.

The CSC theory of tumour initiation and progression was born from early observations that tumour cells differ in their ability to xeno- and auto-transplant (17) and added to by seminal studies of teratocarcinomas (18), small cell lung carcinomas (19) and mammary adenocarcinomas (20). The theory posits that cancer follows the same principles as embryogenesis and normal tissue renewal. Thus, cancer ‘stem’ cells produce all compartments of the tumour much as normal stem cells produce differentiated cells of the adult tissue (24). This model is sometimes seen as controversial, particularly as robust markers for the isolation and purification of CSCs are not available for all malignancies. Nevertheless, stem cell signatures do show prognostic features across multiple tumour types (431–434) and that fact that tumour cells differ in their capacity to xeno- and auto-transplant remains. Moreover, recent breakthroughs in technology development

and in our understanding of stem cell biology could reignite the field. To highlight three such advances: 1) single cell NGS analysis showing stem-cell-like expression patterns in human metastatic breast cancer cells (435); 2) various tissues have been found to contain dormant or quiescent stem cell populations with the ability to regenerate tissues on damage (189–191); 3) certain oncogenic driver mutations have been found to promote the reacquisition of multipotency (71,72). For these reasons and others, Kreso and Dick have put forth a unifying model of CSCs and clonal evolution of genomic clones (24). The authors propose that certain cellular phenotypes within genomic clones may adopt a ‘dormant’ state but be able to reacquire malignancy in a context specific manner. Thus, the CSC and clonal evolution models of tumour development can be reconciled by considering that genomic clones may contain a heterogeneous mixture of cellular phenotypes. Such cellular populations have been described in CRC PDX models (83), where quiescent cellular clones were linked to the acquisition of chemoresistance and later found to express the putative stem cell marker BMI1 (192).

In *Chapter 3*, we identified isogenic cellular clones which gave rise to tamoxifen resistance in MCF7s (*Figures 3.6.2, 3.7.2.2, 3.7.3 & 3.8.1*). Considerable overlap between barcodes enriched in separate replicates was observed, leading us to conclude that an average of 49.6 resistant clonal populations were present in each replicate and that 83% were enriched consistently in more than one cell line. This enrichment process was not the result of genomic clonal selection, however resultant resistant phenotypes did show heterogeneity (for example in 4-OHT sensitivity, ER α function, rapamycin sensitivity and proliferation; *Figures 3.5.3.2, 3.5.4 & 3.5.5*). Moreover, we found evidence that resistant cellular clones may be slow-cycling or even quiescent in nature (*Figures 3.8.2.2 & 3.8.3*). If barcoded clones enriched in tamoxifen resistance were indeed dedifferentiated, this could help explain the heterogeneity we observed in isogenic populations in our system and perhaps even the wider heterogeneity in tamoxifen resistance reported in the

literature (104,277,288,436,437). Supporting this notion, embryonic transcription factors, ER α binding plasticity and (dys)regulation of the epigenome have been inextricably linked to each other and to the development of tamoxifen resistance in the literature (256,285,321). Thus, in this Chapter, we sought to further investigate the role of BC-TICs in our models of tamoxifen resistance.

We begin this Chapter by profiling a selection of ER α + cell lines, the majority of which are derived from MCF7s, for expression of embryonic transcription factors and putative surface markers of BC-TICs. Resistant lines were enriched for *PROCR* (Figures 4.2.1 & 4.2.2) and expression of embryonic transcription factors (*SOX2* & *POU5F1*; but not *KLF4* or *NANOG*; Figures 4.2.1 & 4.2.2). Each of the four assayed transcription factors have been linked to tamoxifen treatment or resistance. We find enrichment of *POU5F1* and *SOX2* in resistant models, each of which has been positively associated with the development of resistance (277,438), *SOX2* overexpression even has a non-significant trend toward poor prognosis in ER α + METABRIC patients (Figure 4.8.2). Conversely, *KLF4* is generally associated with a favourable outcome in breast cancer (280,284) and in our models, it was associated with tamoxifen sensitivity.

Resistant cell lines derived from MCF7-CT were invariably enriched for MFE (Figure 4.2.3) and, perhaps unsurprisingly, we find mammospheres, but not other 3D cultures, are resistant to 4-OHT treatment (Figure 4.2.6). Treatment of first generation mammospheres with 4-OHT also enriches for MFE (Figure 4.2.4). Moreover, we find similar cellular populations are enriched in mammosphere culture (*i.e.* M-ICs) and in tamoxifen resistance (Figure 4.2.8). Together these results suggest that 4-OHT treatment preferentially affects non-M-ICs and will therefore enrich the proportion of M-ICs in a culture. This affect is accompanied by an upregulation of *POU5F1*, *SOX2* and *PROCR* (Figure 4.2.2) suggesting that these genes are preferentially expressed in M-ICs. We see a similar pattern with ALDH activity

assays in PDX models. 4-OHT increases the ALDH⁺ population size (*Figure 4.6.4*) and the ALDH⁺ population is more resistant to 4-OHT (*Figure 4.6.5*).

Our experiments have also shown the obligate need for *POU5F1* expression in 4-OHT treated mammospheres. Whereas SOX2 is required for all mammosphere formation, OCT4 appears to be functionally tied to 4-OHT-selection of BC-TICs in MCF7s (*Figures 4.4.2 & 4.4.3*). Moreover, we find a plausible mechanism by which 4-OHT could trigger OCT4 activity: a validated promotor region in the *POU5F1* gene containing EREs (407), a poor prognosis ER α /FOXA1 binding site (248) and putative tamoxifen-ER α specific binding site (408) (*Figure 4.2.5*). In profiling our OCT4-Vex-Poly MCF7 cell line, we also found evidence of pre-existing 4-OHT resistance (*Figure 4.4.2*) and a reduced necessity for E2-mediated growth (*Figure 4.6.1 & Figure 4.6.2*). Phenotypically, we find OCT4-Vex-Poly to be slow growing and enhanced in BC-TICs /M-ICs by LDA, both of which traits were profiled *in vitro* (*Figures 4.4.2 & 4.4.4*) and *in vivo* (*Figures 4.6.2 & 4.6.4*). Most strikingly, we find OCT4 co-operates with 4-OHT selection to increase BC-TIC activity in the Gold-Standard, *in vivo*, LDA. Together these results hint at a functional role for OCT4 in the development of tamoxifen resistance. Whilst many of the reports in the literature on 4-OHT inducing a BC-TIC-like phenotype can be explained by the observed enrichment for a pre-existing population, OCT4 appears to be functionally required for this enrichment.

Interestingly, in our studies, even though *POU5F1* overexpressing tumours are slow growing and not associated with poor prognosis in most breast cancers, there is a significant association with worse overall survival in ER α ⁺ patients who are currently undergoing hormonal therapy (*Figure 4.8.3*). In the METABRIC cohort, we also find *KLF4* expression enriched in ER α ⁺ patients and in ER α ⁺ patients treated with hormone therapy, by Kaplan-Myer, we see a non-significant survival benefit in those patients with upregulated *KLF4* (*Figures 4.8.1 & 4.8.2*). NANOG has recently

been shown to drive tumorigenesis in TNBC by stabilising with Cx26 and activating FAK (429). In our analysis, however, NANOG is neither associated with the development of resistance or enrichment in the TNBC rich Integrative Clusters (IntClusters) such as IC10 (*Figures 4.8.1 & 4.8.2*).

Recently, *Bhatt et al.*, characterised the role of OCT4 as a novel ER α binding partner involved in a tamoxifen specific ER α binding program and ultimately tamoxifen resistance (438). The authors performed motif analysis of tamoxifen specific ER α binding sites identified by *Welboren et al.*, in MCF7s (408) and found an enrichment for Nkx3-1 and Oct-transcription factor homodimer motifs. The finding was subsequently confirmed by ChIP assays for OCT4, demonstrating its specific recruitment to regulatory sites of genes induced by tamoxifen, but not oestradiol, bound ER α . Interestingly, further experiments found *POU5F1* expression to be basally repressed by Nkx3-1 in MCF-7 cells, and that 4-OHT was able to elevate Nkx3-1 degradation through a p38MAPK-dependent phosphorylation of the E3 ligase, Skp2. The authors conclude by presenting evidence that OCT4 participates in ER α transcriptional complexes along with p38MAPK and Skp2 in a tamoxifen-dependent manner, and that this activity leads to activation of a 4-OHT-transcriptional program and proliferation of the TamR cell line (438).

In another study, *Cho et al.*, found downregulation of carboxy terminus of HSP70-interacting protein (CHIP) E3 ubiquitin ligase in MCF7s and MDA-MB-231s cultured as mammospheres (439). The authors found that CHIP depletion increased mammosphere formation, whereas over-expression reduced formation. By mass spectrometry interactome reconstruction, the authors identified OCT4 as a direct binding partner of CHIP, and later functional studies found CHIP over-expression decreased OCT4 stability through ubiquitination and proteasomal degradation, together with decreased proliferation and CSC-like side-population size (439). The authors conclude that CHIP-induced OCT4 ubiquitination is important in breast CSCs and suggest regulation of CHIP expression and/or OCT4 stability as a potential

future therapeutic strategy for ER α + breast cancer (439). Together, these studies serve to corroborate the crucial role of OCT4 for M-IC maintenance and link to 4-OHT resistance described in this Chapter.

Chapter 5

Lineage Tracing of PDX models with High Complexity Barcodes

Aims of the chapter

- 1) Optimise lentiviral infection protocols to genetically modify breast cancer PDXs.
- 2) Ensure genomic stability of PDX models on lentiviral infection.
- 3) Passage and expand barcoded PDX models.
- 4) Interrogate clonal repopulation dynamics by NGS of CT barcoded PDX models.
- 5) Investigate clonal selection in the PDX metastatic setting.

5.1 General Introduction

Despite remarkable advances in our understanding of the progression of human malignancies and the molecular events that underpin tumour survival, new therapies often fail to show significant efficacy in clinical trials. It could be argued that clinical trials fail to sufficiently stratify patients based on relevant biomarkers of drug response: the response rate of an unscreened population to a molecularly targeted therapy typically lies between 10 and 20% (339). Patient stratification based on molecular determinants of drug efficacy and tumour heterogeneity allows for significantly greater responses - exemplified by the success of ALK kinase inhibitors in EML4-ALK positive non-small cell lung cancers (440). However, even with patient stratification, clinical responses can be fleeting, often adding only 6-12 months before disease progression (339). We have argued previously that these limitations are driven, at least in part, by extensive reliance on preclinical models that fail to accurately reflect tumour heterogeneity (23,289). In order to halt unsustainable rates of attrition in the drug discovery process, and to develop cancer medicines giving lasting patient benefit, we must develop a new generation of preclinical models capable of reflecting the complex heterogeneity found in human cancers. PDX models prevail as arguably the most powerful in this regard; because they capture multiple aspects of tumour heterogeneity (195,196,198).

By capturing intra- as well as intertumour heterogeneity, PDX models have a clear advantage over traditional models, supporting their use in oncological drug discovery and preclinical development. PDX models recapitulate cell-autonomous drivers of heterogeneity: exhibiting genomic clonal dynamics reminiscent of their originating tumour sample (195,197). For example, *Eirew et al.*, reconstructed the genomic clonal dynamics of a panel of breast cancer PDX models using PyClone (195). In each of the 15 cases examined, clonal diversity was reduced by xenotransplantation. This varied from extreme engraftment bias, selecting minor clones present in the sample of origin, to only moderate clonal selection. In a separate study, *Ding et al.*, found PDX models established

from a basal like breast cancer were more representative of the patient's metastatic lesion than primary tumour (153). Together, these observations suggest that deterministic, fitness-based mechanisms underline tumour-dependent clonal selection observed on engraftment. *Eirew et al.*, further observed variable clonal dynamics between PDXs established from different molecular subtypes but similar dynamics in parallel xenografts established from the same sample, underscoring the need for better representation of tumour molecular subtypes (7,195).

Moreover, phenotypically distinct isogenic cellular clones have been shown to drive resistance to chemotherapy in colorectal PDX models (196). In an elegant study by *Kreso et al.*, it was found that minor 'Type IV' subclones in colorectal PDXs were able to repopulate the tumour bulk after treatment with chemotherapy (196). These quiescent cell populations were genetically similar to their highly proliferative counterparts, and were later linked to the BMI1+ population thought to act as reserve stem cells of the intestinal and colonic crypts (192). As the cellular clones defined in this study were isogenic, their phenotype may have been driven by microenvironmental cues capable of modulating cellular transition between distinct gene expression patterns or epigenetic attractor states (441). Hence, composition of the TME could profoundly alter both a cells propensity to malignancy and the heterogeneity we hope to preserve in PDX models, underscoring many of the limitations of this model.

Aside from clonal dynamics driven by intrinsic differences in a cell's genetic or epigenetic background, intratumour heterogeneity can be influenced by tumour-extrinsic factors in the non-cell-autonomous compartment (23). As cancers develop, tight regulation of the ECM is lost and tissue architecture begins to degrade (125). ECM driven oscillations between signalling pathways such as those described could have profound effects on propensity to malignancy. Furthermore, solid state ECM interactions are necessary for cells to maintain stem cell properties and a regulated ECM helps maintain the stem cell niche (126). As cancer is often associated with a blurring of the boundaries between stem

and differentiated cells, it is possible that a loss of structured ECM is essential for the stability of multiple sub-dominant cellular clones within a tumour (124). The TME is further characterised by an influx of stromal cells. Infiltrating cancer associated fibroblasts (CAFs) can often confer resistance to cytotoxic and targeted therapies (442), however, recent studies confer on fibroblasts a degree of plasticity, with anti-tumour properties observed in some populations (323). Due to the high levels of CAF infiltrates seen in some tumour types, heterogeneity within their population would undoubtedly confer differential properties to the tumour bulk. We, and others, have found that human stromal cells are gradually replaced by murine equivalents upon engraftment in the mouse, suggesting that implanted human cells retain the ability to recruit murine accessory cells to their niche (23,198). However, it should be noted that some differences exist between ligand repertoires of human and murine fibroblasts (443) and more research is required to properly understand how faithfully the TME is recapitulated in the murine host.

Finally, considering the crucial role of the immune system in tumour progression, perhaps the most obvious limitation of PDX models is the necessity for severely immunodeficient host animals (23). The pro-inflammatory microenvironment established by CD8⁺ T cells, M1 polarised TAMs, NK cells and others can lead to the recruitment of numerous immune suppressive components inhibitors (23,43,444). TAMs and myeloid derived suppressor cells have been implicated in resistance to anti-angiogenic therapy. Additionally, macrophage and CD4⁺ T cell recruitment following intensive chemotherapy in breast cancer patients is associated with significantly reduced recurrence free survival (444). Clearly, the future co-engraftment of immune components into PDX models would facilitate both the study of novel therapies targeting tumour-immune interactions and allow for basic research into patient specific cross-talk between tumour progression and immune surveillance.

Recently, we have described a biobank of breast cancer PDX models with extensive molecular, genomic and clinical annotation (198). Our study adds to a growing body of evidence that PDX models reflect the complex molecular and architectural features of their originating patient tumour samples (23,307,445,446). Moreover, we developed an integrated platform allowing the use of our PDXs in high throughput combinatorial drug screening. Our work, together with other preclinical studies, for example, a recently published drug screen in 1,000 PDX models, provided crucial evidence supporting the future utility PDX models in predicting human clinical trial drug responses (198,447).

In summary, PDX models can partially recapitulate the complex population dynamics of a human malignancy. Reproducible patterns of clonal dynamics suggest engraftment bias may represent a non-stochastic selection event, which defines a PDX model rather than limits its utility. However, many limitations of this model exist, and as with all model systems, we must be careful to interpret our results within the context of the limitations of our tools.

In this chapter, we describe efforts to extend our lentiviral clonal tracing technology to our breast cancer PDX cohort. The ability to deconvolute cellular and genomic clonal dynamics in these highly complex models of human malignancy offers the potential to redefine our understanding of tumour heterogeneity (289). Our models have established utility in studying the mechanisms of drug response and resistance, and this new development could extend their utility as a powerful tool to unravel the clonal dynamics in response to therapeutic perturbation and the role of tumour heterogeneity in drug response.

5.2 Lentiviral Infection of PDX Models

We sought to track the fate of individual cellular clones through repopulation of multiple rounds of PDX model passage. It was reasoned that a heritable molecular barcoding strategy, similar to those detailed in *Chapter 3*, would allow an assessment of the relative growth dynamics and fitness of individual cell in our PDX biobank. We could then introduce selective pressures, such as chemotherapeutic treatment regimen, to assess the relative contribution of each cellular clone to clinically meaningful phenotypes, such as the development of drug resistance and the processes associated with metastasis. This molecular barcoding technique would require high resolution labelling of $\sim 1 \times 10^6$ PDX cells in order to begin capturing heterogeneity on the scale present in human malignancies. The ClonTracer library, established and characterised by *Bhang et al.*, features a 30-nucleotide degenerate-sequence and theoretical complexity of $\sim 72 \times 10^6$ (334). Hence, this system was again selected for these studies. The ClonTracer library has been used to individually label $> 1 \times 10^6$ cells and assess their relative contributions to drug resistance in models of CRC (334) and breast cancer (*Chapter 3*), but has, to our knowledge, never been applied to track the fate of cellular clones in PDX models.

We sought to introduce ClonTracer degenerate barcodes into our PDX models by lentiviral infection, using a similar process as developed for MCF7s during our tamoxifen resistance studies, detailed in *Chapter 3*. Complexity calculations are detailed extensively in *Chapter 3 (Section 3.4)* and *Chapter 2 (Methods, Section 2.3)*. Briefly, we have shown highly significant concordance between calculated number of plasmid molecules and unique barcode sequences by NGS *Figure 3.4.2*, and that in a sampling of 1×10^6 barcodes, 96% are unique, *Figure 3.4.5*. We have also shown no PCR bias in a random sampling of 24 individual barcode clones (*Figure 3.4.2*) and have successfully used the technology to label and track a highly diverse population of MCF7s during the development of resistance to 4-OHT (*Figures 3.6.1, 3.6.2 & 3.7.2*). Here, we sought to use the same ClonTracer lentiviral library to characterise the repopulation dynamics in our breast

cancer PDX models. It is hoped that characterisation of this initial model will allow resistance studies in our highly annotated PDX biobank.

Using the same lentiviral preparation as produced and reported in *Figure 3.4* we infected *in vitro* PDTC cultures of a HER2- ER α - breast cancer PDX model, STG282. Several PDTC cultures tested were found to be resistant to puromycin at concentrations up to 10,000ng/ml. In *Figure 5.2.1 A* we show resistance in STG282-x3 at concentrations up to 5,000ng/ml. The ClonTracer insert contains a red florescent protein (RFP) reporter alongside a puromycin resistance cassette, to minimise any potential off target effects of puromycin in these cells, we opted for flow cytometry based viral titration. *Figure 5.2.1 B-F* shows the gating strategy used when selecting RFP expressing (RFP+) cell populations. First, we selected a population based on forward (size) and side (granularity) scatters to likely minimise debris collection (*Figure 5.2.1 B*), doublet discrimination by both forward scatter (FSC), *Figure 5.2.1 C*, and side scatter (SSC), *Figure 5.2.1 D*, was then performed to ensure single cells in downstream analyses. Doublets have been shown to adversely affect the purity of cell sorted populations (448), and our PDTC models grow as organoid-like structures *ex vivo* (198), hence we used both forward and side scatter exclusion to ensure as pure an RFP+ population as possible. Next, we excluded dead cells from our analysis using DAPI as a viability marker (*Figure 5.2.1 E*), and selected only human cells, using a marker of mouse major histocompatibility complex (MHC) pan MHC-I/MHC-II (H-2k^b/H2-D^b; *Figure 5.2.1 F*). Compensation for fluorescent bleed in this multicolour flow cytometry experiment was carried out in FlowJo (version 10.5.0) by analysing single stained controls accompanying each experiment. The panels shown in *Figure 5.2.1* are representative of the majority of our flow experiments; PDTCs tend to be high in doublets and 25-30% comprised of murine cells. H-2k^b/H2-D^b has been validated in our lab as 93% sensitivity and 92% specificity for detecting host stromal cells in PDX models by using mouse whole blood as a positive control (198).

We first optimised our lentiviral infection techniques in PDTcs using HIV-ZsGreen1, in order to preserve stocks of a validated ClonTracer lentiviral cocktail. *Figure 5.2.2 A* shows infection of STG282-x3 with 1×10^{10} TU/ml of Lenti-HIV-ZsGreen1 as determined by functional titre in HEK293s (*see; Methods, Section 2.3.3*). Only 8.22% of viable human single cells were successfully infected with this relatively high viral titre. We reasoned that such a low efficiency of selection could indicate that only a subpopulation of PDTcs were susceptible to lentivirus, however, HIV-ZsGreen1 positive cells seem to be relatively evenly distributed across forward and side scatter distributions (*Figure 5.2.2 B*). Certain primary human cells, particularly those that are quiescent in nature, are reproducibly challenging to transduce with lentiviral constructs (449,450). Clearly, quiescent (*e.g. Type IV clones (196)*) could hold particular importance in our later functional studies (451), and a molecular barcoding technique which selected for non-quiescent cells should be avoided.

ZsGreen1+ cells were isolated from the bulk ZsGreen1- STG282-x3 population by fluorescence activated cell sorting (FACS) and implanted into the flank of NSG mice as described in *Methods, Section 2.1*. Implanted regions were palpated periodically for signs of tumour growth, and estimated tumour volume recorded by the BRU. Interestingly, implanted ZsGreen1+ STG282 cells were markedly slower growing when compared to models of the same passage (*Figure 5.2.2 C*). Ninety-five days after implantation, ZsGreen1+ tumours were only beginning to become palpable; STG28-x3s, on average, were 800mm³ by this point. However, after 100 days, ZsGreen1+ tumours grew at a rate not dissimilar to parental STG282s. After 140 days, tumours had grown to 900mm³ and so animals were culled and PDX tumour fragments re-implanted into four secondary NSGs. Each resultant STG282-x3 HIV-ZsGreen1 x2 tumour grew at a rate not dissimilar to wild type STG282-x3 (*Figure 5.2.2 D*).

We reasoned delayed growth on implantation could, thus, be at least partially attributed to the process of flow sorting rather than selective infection of a slow growing sub-

population of cells. Although quiescent cells have been shown as less likely to be successfully transduced by lentivirus, slow growth of our infected xenograft models due to the potential physical stress associated with flow sorting would mask this effect.

Next, we went on to infect STG282-x3 cells with the same ClonTracer lentiviral library as developed and validated in *Chapter 3 (Section 3.4)* and *Chapter 2 (Methods, Section 2.3)*. 1×10^7 PDTCs were isolated from frozen stocks of STG282x3 and cultured *ex vivo* in the presence of ClonTracer lentiviral construct. After 24hrs, PDTCs were screened for RFP+ cells by flow cytometry. *Figure 5.2.3 A* shows the gating strategy for isolation of RFP+ cells following our live human singlet gating process detailed in *Figure 5.2.1*. A lentiviral dose of 1.2×10^{11} TU/ml resulted in 9.34% RFP+ cells by flow cytometry, suggesting that the majority of cells would be individually labelled (the result of a single integration event (360)). As with previous viral constructs, integration events were spread across a spectrum of forward and side scatter discriminated particles (*Figure 5.2.3 B*). The resultant “STG282CT” cells (9×10^5) were isolated by FACS and implanted in the flank of an NSG mouse, following the same experimental procedures as with HIV-ZsGreen1 infected cells, shown in *Figure 5.2.2*. Viability staining by trypan blue exclusion (348) following flow sorting but prior to implantation showed viability at 64%. As before, infected cells resulted in slower growing tumours on first implantation, but normal growth rates were restored on passage (*Figure 5.2.3 C*). STG282CT-x0 was harvested after 135 days, dissociated into single cells and re-implanted into 10 NSG mice in an attempt to increase barcode library representation. A sample was also taken for histology, with staining by H&E, together with IHC for EpCam and ECadherin, shown in *Figure 5.2.5*. Heterogeneity in protein expression of each marker is seen to be maintained from parental STG282x3 to the STG282CT-x0 derivative.

Finally, we profiled STG282CT by shallow whole genome sequencing (sWGS) to identify copy number variations associated with lentiviral integration hotspots, and/or a loss of heterogeneity associated with selective pressures from lentiviral infection itself, or the

subsequent flow sorting process. Visualisation is provided through the QDNaseq Bioconductor package. *Figure 5.2.5 A* shows the originating STG282 metastatic tumour sample, *Figure 5.2.5 B* a representative example of STG282CT-x4 and *Figure 5.2.5 C* STG282CT-x0. No overt differences were observed by visual or quantitative analysis of the copy number profiles of infected STG282CT and either the originating sample or a long term PDX culture. The degree to which PDX models preserve the genomic heterogeneity of their originating sample is covered extensively in *Bruna et al.*, (198).

Figure 5.2.1 – Puromycin resistance & PDTC gating strategy

A) Shows resistance to puromycin in STG282 (up to 5,000ng/ml), presented as viability versus non-treated control (NT) by Cell-TitreGlo viability assay. **B-F)** show our gating strategy by flow cytometry to isolate individual, human, live cells carrying a lentiviral construct of interest. Specifically, **B)** A population with suitable forward (size) and side (granularity) scatter parameters to minimise cellular debris, to likely minimise debris. **C)** doublet discrimination by forward (FSC) and **D)** Side (SCC) scatter. **E)** Dead cells are Excluded by DAPI and **F)** murine stromal cells were excluded by pan mouse MHC-I/MHC-II (H-2kb/H2-Db). Multicolour flow cytometry compensation was carried out using the recommended manufacturers procedures and single stained samples (DAPI positive control was achieved by boiling a suspension for 20mins to reduce cell viability). Cells were analysed and sorted using the BD FACSAria or the BD Influx FACS machine with data files analysed using FlowJo 10.5.0 (for MacOS). The CI Flow Cytometry Core aided in the completion of these experiments. **Figure Overleaf.**

Figure 5.2.1 – Puromycin resistance & PDTC gating strategy

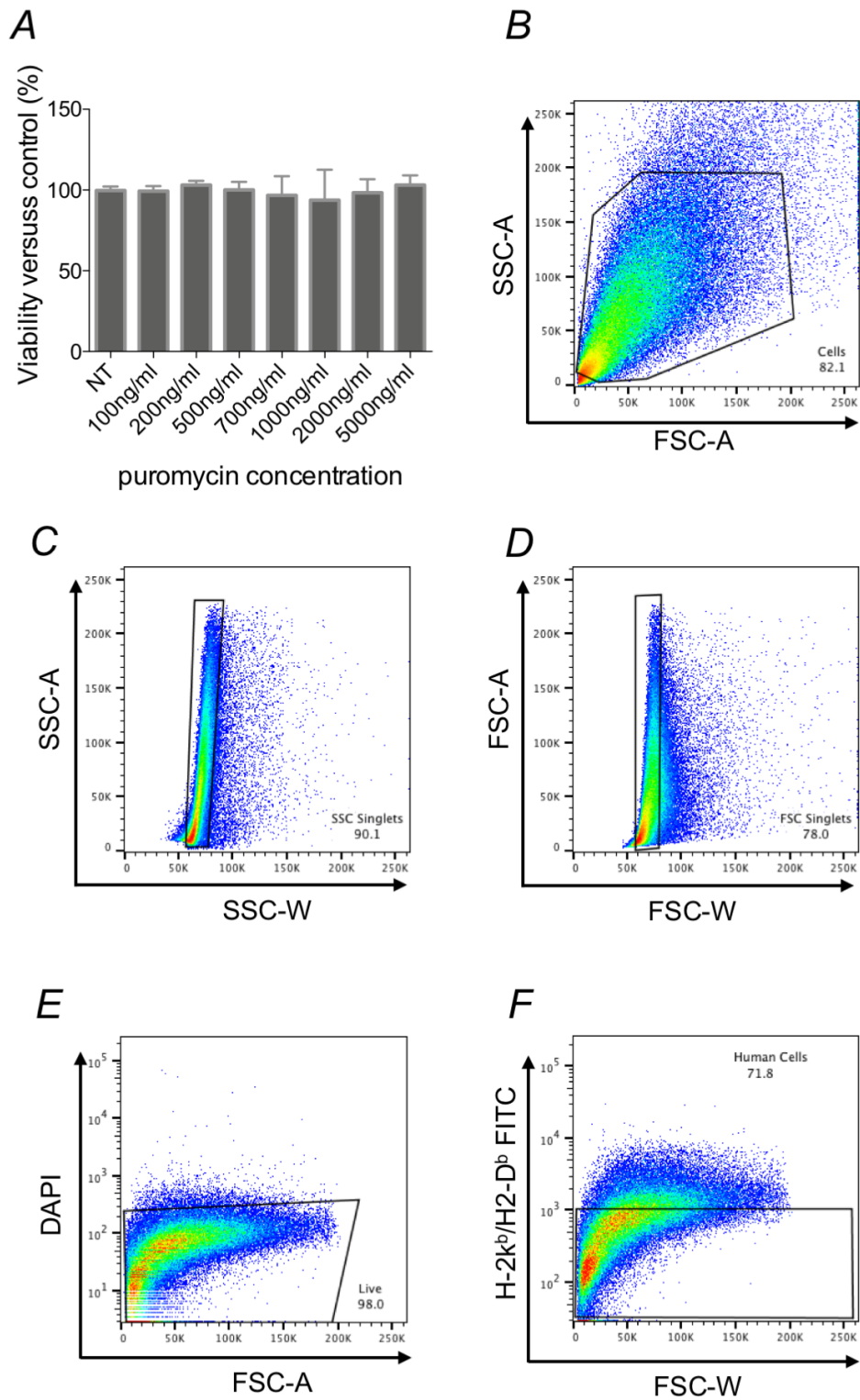


Figure 5.2.2 – STG282x3 HIV-ZsGreen1 implantation

A) STG282-x3 were cultured in the presence of 1×10^{10} TU/ml of Lenti-HIV-ZsGreen1 as determined by functional titre in HEK293-Ts (see; Methods, Section 2.3.3) in conditions optimised to encourage lentiviral infection. 8.22% of viable human single cells were ZsGreen1 positive after 24hrs infection. **B)** STG282CT-HIV-ZsGreen1+ cells seemed to be relatively evenly distributed across forward and side scatter distributions. **C)** ZsGreen1+ cells were isolated by FACS and implanted into the flanks of NSG mice. Tumour size was estimated by palpation periodically and animals maintained by BRU staff. After 140 days, tumours had grown to 900mm^3 and so animals were culled and STG282CT-HIV-ZsGreen1 re-implanted into four secondary NSGs. **D)** By passage x2, infected tumours grew at a rate similar to parental models. **Figure Overleaf.**

Figure 5.2.2 – STG282x3 HIV-ZsGreen1 implantation

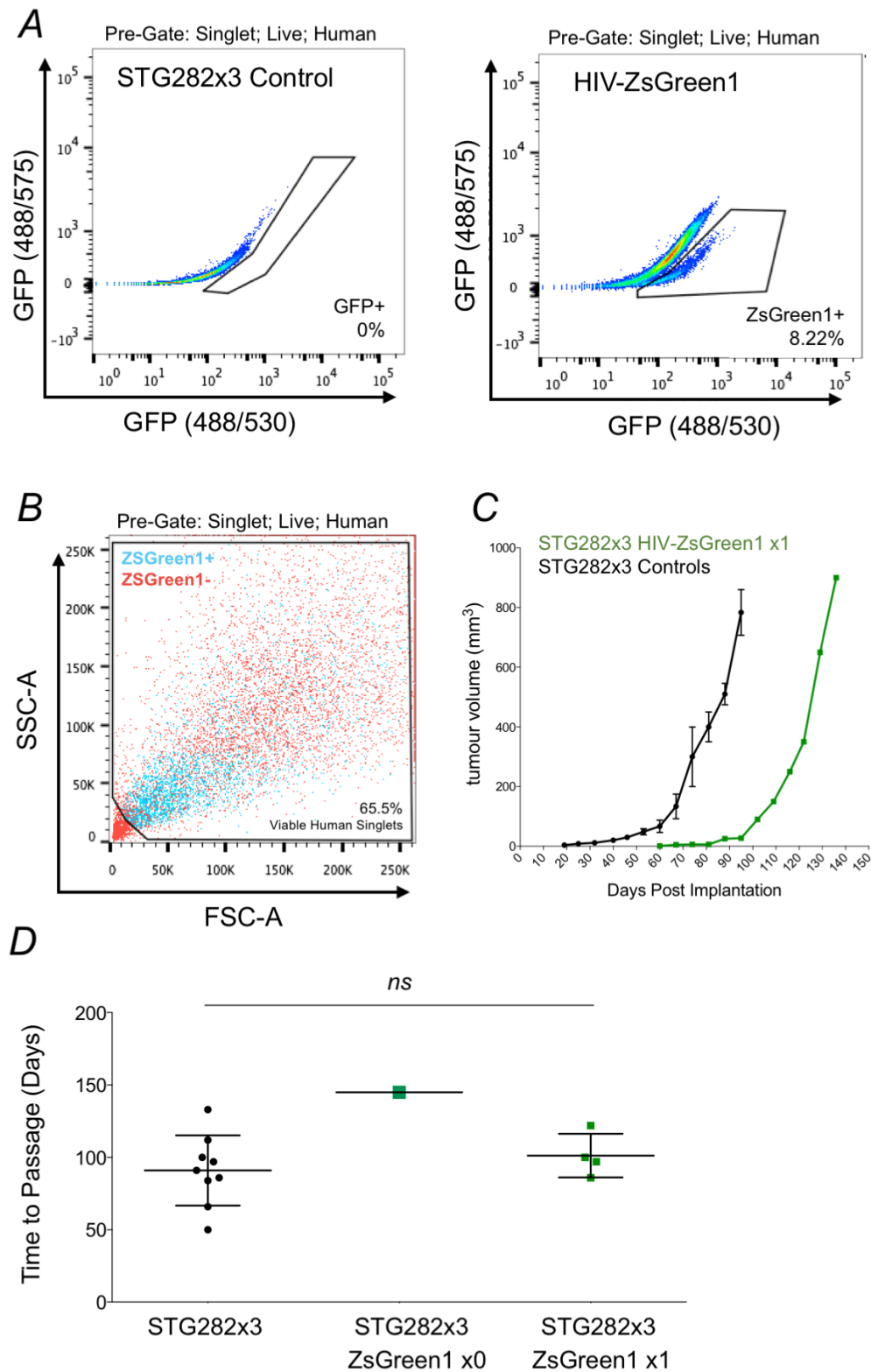


Figure 5.2.3 – STG282x3 ClonTracer infection & implantation

*STG282x3 cells were infected with the same ClonTracer lentiviral library developed and validated in Chapter 3 (Section 3.4) and Chapter 2 (Methods, Section 2.3). **A)** 1×10^7 PDTs were isolated from frozen stocks of STG282x3 and cultured ex vivo in the presence of ClonTracer lentiviral construct (1.2×10^{11} TU/ml). After 24hrs, PDTs were screened for RFP+ cells by flow cytometry, we found 9.34% RFP+. This MOI suggests singular integration events in the majority of cells (360). **B)** Integration events were spread across a spectrum of forward and side scatter discriminated particles. **C)** The resultant “STG282CT” cells (9×10^5) were isolated by FACS and implanted in the flank of an NSG mouse. Viability staining by trypan blue exclusion (348) following flow sorting but prior to implantation showed viability at 64%. As before, infected cells resulted in slower growing tumours on first implantation, but normal growth rates were restored on passage. The CI BRU aided in the running and analysis of these experiments. **Figure Overleaf.***

Figure 5.2.3 - STG282x3 ClonTracer infection & implantation

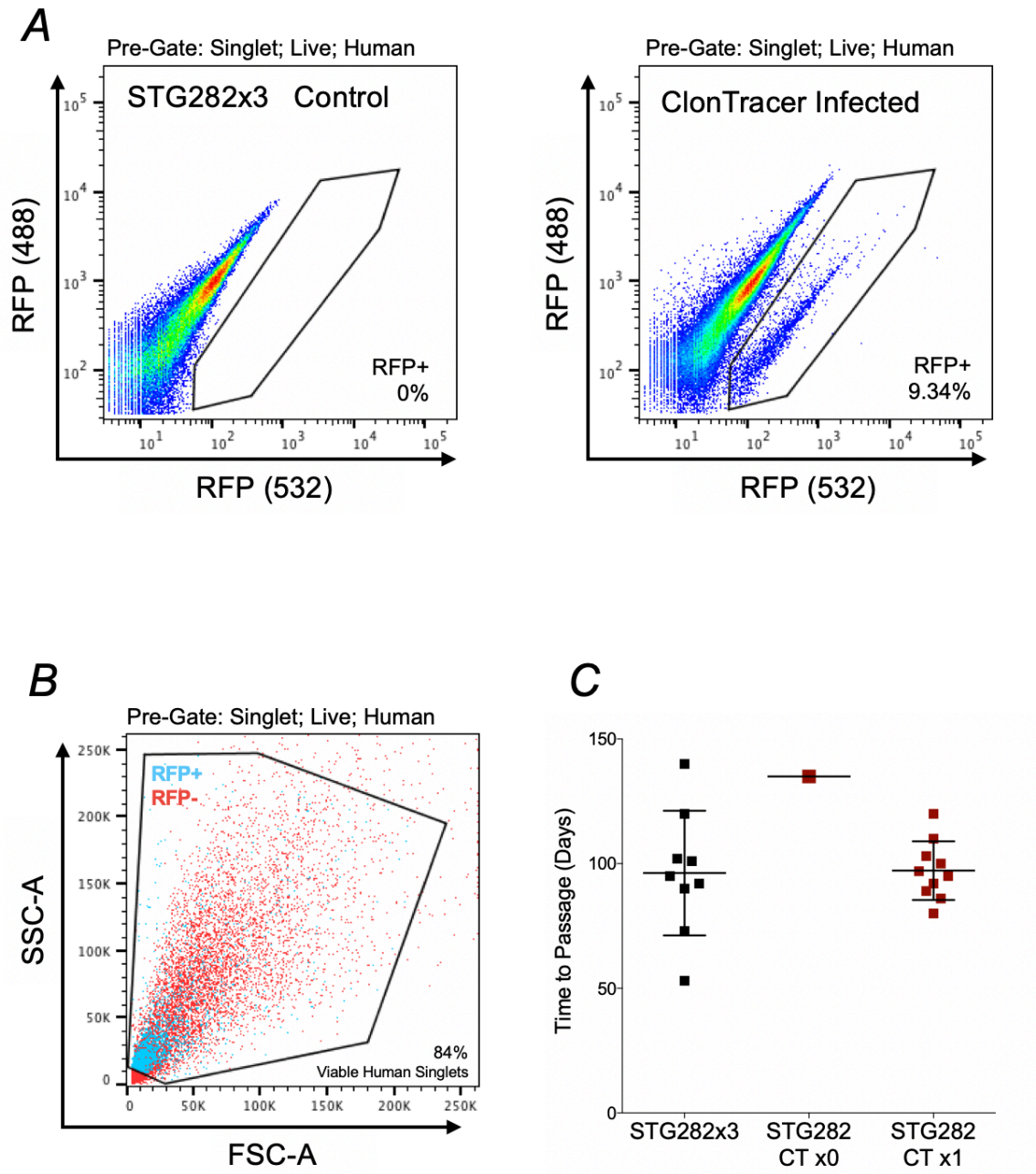


Figure 5.2.4 – STG282x3 ClonTracer x0 (STC282CT-x0) histology

*Representative images of samples STG282CT-x0. Staining shows H&E and EpCam & ECadherin IHC. Comparison to parental lines shows maintained heterogeneity in protein expression of each marker. Dr H Raza Ali and the CI Histopathology Core aided in the running and analysis of these experiments. **Figure Overleaf.***

Figure 5.2.4 - STG282x3 ClonTracer x0 (STG282CT-x0) histology

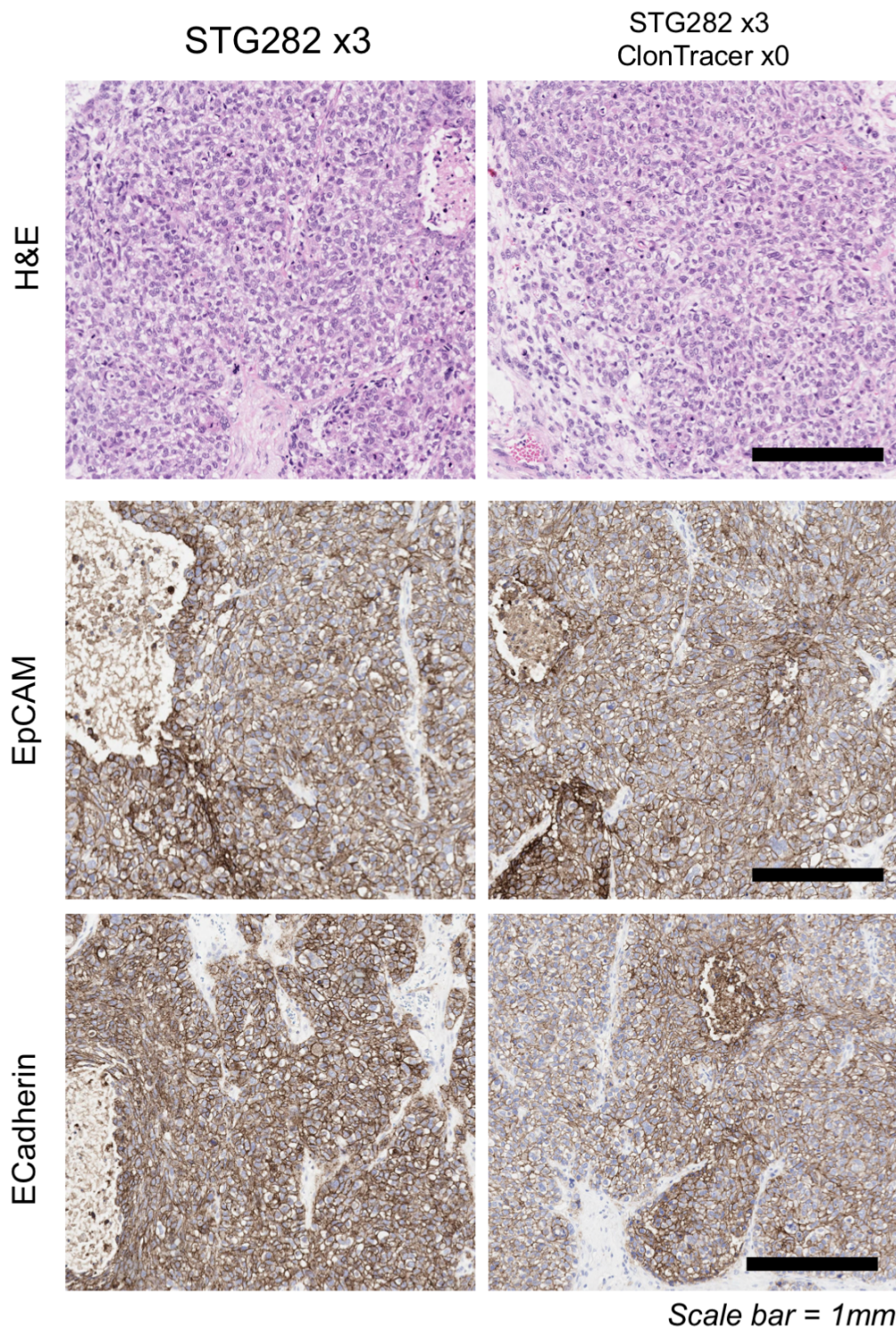
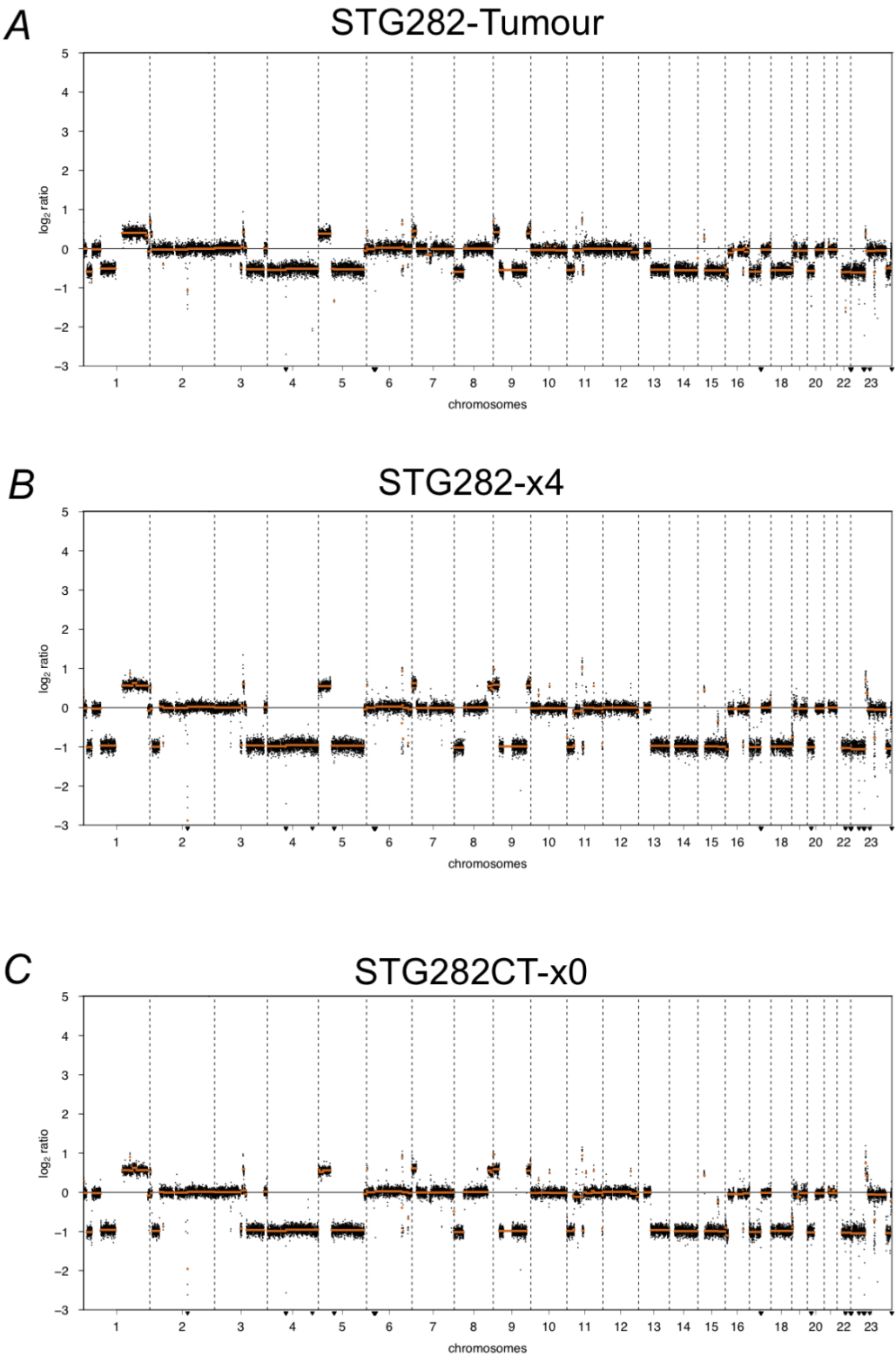


Figure 5.2.5 – STG282CT-x0 sWGS

*STG282CT was profiled by sWGS to identify copy number variations associated with lentiviral integration hotspots, and/or a loss of heterogeneity associated with selective pressures from lentiviral infection itself, or the subsequent flow sorting process. Visualisation is provided through the QDNaseq Bioconductor package. Analysis was carried out by Dr Oscar Rueda. **A)** shows the originating STG282 metastatic tumour sample, **B)** a representative example of STG282CT-x4 and **C)** STG282CT-x0. No overt differences were observed by visual or quantitative analysis of the copy number profiles of infected STG282CT and either the originating sample or a long term PDX culture. **Figure Overleaf.***

Figure 5.2.5 - STG282CT-x0 sWGS



5.3 Loss of Diversity in STG282CT PDX Models

Having established a barcoded PDX model, STG282CT, we sought to expand this model across multiple passages. We reasoned that by analysing barcode presence at different passages and in different arms we could study clonal selection and competition in our PDX models, and that by analysing their relative abundance we could derive a model of clonal fitness. Moreover, the ability to trace the fate of individual barcoded cells in this, and future, PDX models, could give us insight into the clonal origins of drug resistance and metastasis. *Figure 5.3.1* shows a schematic of the expansion of the STG282CT model. Model numbers in red represent those where NGS and barcode analysis was performed, and red triangles indicate models in which micro or local metastases were found.

After lentiviral infection, flow sorting and implantation (*Figure 5.2.3*), we estimated labelling of approximately 900,000 individual PDTCs. Viability staining by trypan blue exclusion (348) found 36% of cells were non-viable, suggesting ~600,000 individually labelled PDTCs. A typical PDX fragment (for reimplantation or PDTC dissociation) contains between 1×10^6 and 5×10^6 viable cells (198), with each tumour containing 1×10^7 to 5×10^7 cells. Assuming complete mixing of barcoded populations, together with equal growth propensity and clone forming ability, we calculated that reimplantation of STG282CT-x0 into 10 NSG hosts, followed by tumour expansion and PDTC harvesting, would result in >150-fold library representation. PDTCs from this STG282CT-x1 pool could then be mixed and re-implanted into NSG to allow equal barcode distribution in passage x2 models. We hoped to validate this library expansion and representation so that future drug resistance lines could directly compare the fate of cellular clones in treatment and control arms of a drug resistance study, similar to that designed in *Chapter 3*. *Figure 5.3.2 A* shows our STG282CT expansion and pooling strategy to increase barcode library representation.

ClonTracer barcode regions were amplified from genomic DNA by PCR and analysed by NGS as detailed in *Chapter 2, Methods 2.5*. Briefly, Barcode sequences were extracted from NGS data after standard QC and demultiplexing with FastX, and analysed using the ClonTracer software package (135). Counts, frequency distributions and barcode abundances were then processed in R. *Figure 5.3.2 A* shows a pooling strategy to check library complexity in STG282CT-x1 models. Despite a theoretical complexity of 900,000 uniquely marked cells, with up to 500-fold representation, we found only 1,034 unique barcode sequences in a sampling of mixed cells from 10 STG282CT-x1 models (STG282CT-x1_mix; *Figure 5.3.2 B*). In total, 1×10^6 cells from this pool were sequenced, to a depth of approximately 120×10^6 reads, with resultant analysis showing approximately 57×10^6 barcode sequences falling into 1,034 unique sequences (clones).

Of the 10 progeny tumours from this mix (STG282CT-x2_1 through STG282CT-x2_10), three were sequenced for barcode analysis. These second-generation tumours harboured a total of 1,231 unique barcode sequences, of which 578 were unique to passage x2 models and 653 were also seen in STG282-x1 (*Figure 5.3.2 B*). With >50% of barcodes from a deeply sequenced population of cells (STG282CT-x1) potentially going undetected, we sought to better understand clonal growth patterns. *Figures 5.3.2 C, D & E* show the relative abundance of the top 100 unique barcodes, *i.e.* the fraction of usable reads in each sample represented by each of 100 unique barcodes across the three samples. In each pairwise comparison, there is significant overlap in barcode abundance ($p < 0.001$), though the correlation is far from perfect ($r^2 = 0.35, 0.11$ & 0.41 for *panels C, D & E* respectively).

In STG282CT-x2_8 & STG282CT-x2_4, for example, abundance of unique barcodes ranged from 1.1×10^{-4} to 0.8×10^{-1} and 1.3×10^{-4} to 1.4×10^{-1} , meaning that a detected barcode could be present in as few as 0.01% or as many as 14% cells. Moreover, the same barcode could be detected in 10% of cells in one sample (in this case STG282CT-x2_4) and as few as 0.02% in another (STG282CT-x2_8). This wide variability in barcode

abundance between samples led us to consider that asymmetric growth patterns of cells in PDX models may ultimately preclude an exact matching of barcodes between a control and treatment arm, as we had developed for MCF7s (see: *Figure 3.5.1*).

In an elegant study by *Nguyen et al.*, known numbers of single-barcoded control cells were sequenced alongside barcoded PDX and cell line xenograft models in order to estimate cell number from relative barcode abundance (197). We sought to further understand the dynamics of clonal repopulation in our models by following a similar protocol. Hence, we prepared clonally derived (*i.e.* single barcode sequence) ClonTracer constructs prepared by *E. coli* transfection, prepared lentiviral constructs in HEK293-Ts and transduced fresh HEK293 cell preparations at <10% MOI (0.5-1.5 x10⁹ TU/ml concentrated viral titres for each individual barcode). After puromycin selection, known concentrations of single cell HEK293-ClonTracer cell suspensions were spiked into STG282CT samples prior to DNA extraction and NGS barcode sequencing. See *Methods, Sections 2.3 & 2.5.4*.

Across 24 sequenced PDX samples, barcodes spiked in at 500 or more cells were detectable in 100% of cases, and a good correlation was observed between the number of cells input and relative barcode abundance ($r^2 = 0.998$, $p < 0.0001$; *Figure 5.3.3 A*). Below 500 cells, the correlation between input cells and barcode abundance was markedly less clear, and our ability to detect clones fell to 60%. Of note, the strong correlation between input cell number and relative barcode abundance enabled us to derive an equation relating the two variables (*insert; Figure 5.3.3 A*). We used this relationship to calculate estimated cell number for spike in controls (*Figure 5.3.3 B*) and found good concordance with the known number of spike-in cells, across a wide range of values (~1,000 to 50,000 cells; $r^2 = 0.998$, $p < 0.0001$), though the estimated clone size fell markedly below actual clone size at 500 cells or below.

This analysis reinforced the need for a threshold of confident detection and enables us to relate relative barcode abundance to actual clone size in future studies. We set a barcode abundance threshold of $\sim 4 \times 10^{-6}$ (equivalent to 500 cells), as our limit of confident detection. This limit was slightly variable in terms of RBA in each model, but always equivalent to 500 cells, offering a useful internal control to directly compare clonal size in matched replicate implantations.

Aside from estimating the absolute size of a clone, in terms of cell number, this relationship allowed us to estimate the clone-initiating cell (CIC) frequency, *i.e.* the proportion of input cells of a given clone, able to establish a detectable clone on transplantation. In all, we found remarkably fewer clones than input barcode numbers. Across 26 sequenced samples, we find a total of 2,637 unique barcodes, suggesting that as few as 1 / 189 cells are able to form clones. Though the complexity in terms of unique barcode count was lower than originally expected, this figure is relatively concordant with a study by Nguyen and colleagues, who used barcoded PDX and cell line xenograft models to study tumour initiation (197). The authors report decreasing CIC frequency values over more than two orders of magnitude (from 1 / 7 to 1 / 3,000 of cells transplanted in one model, and 1 / 160 to 1 / 12,500 in another) as an inverse function of the input cell number (197).

The above analyses led us to question the possibility of universally increasing barcode representation for drug resistance screens, by expansion in separate animals followed by cell pooling. We sought to understand how clonal diversity increased or decreased across our models. Diversity indexes are often used in ecology to assign a quantitative measure that reflects how many different species can be found in a dataset or community, whilst simultaneously considering how evenly the basic entities (such as individuals) are distributed among those species. We computed one such measure, the SDI (H'), to quantify diversity amongst cellular clones in our PDX models (204,363). *Figure 5.3.4* shows cumulative RBA as a function of the number of unique barcodes

counted, alongside the SDI for each of three parallel passage trees from our PDX expansion map (*Figure 5.3.4*). In each case, diversity is seen to decrease on passage. Not surprisingly, the earliest passage PDX (STG282-x1) was comprised of a diverse group of relatively low frequency clones (showing a lower running total of barcode abundance as the number of unique barcodes increases). Moreover, with an SDI of 324, this is by far the most diverse sample analysed. In each cumulative RBA plot, the corresponding STG282-x2 sample shows early dominance, indicating the presence of a handful of highly abundant barcodes, in each case, the SDI is similarly markedly lower for STG282-x2 than STG282-x1. In *Figure 5.3.4 A*, passage x2 and x3 show relatively similar SDIs, with both passage x4 models actually nearly 1.5-fold more diverse than either. Indeed, although this arm shows an overall trend for decreased diversity with passage, the association is not significant ($r^2 = 0.3159$; $p = 0.0908$). The arm depicted in *Figure 5.3.4 B* is most clear in the trend of decreasing diversity over time ($r^2 = 0.9411$; $p < 0.0001$). Whereas in *Figure 5.3.4 C*, we see similar SDIs in both passage x2 and x3 ($r^2 = 0.6982$; $p = 0.078$). Overall these analyses demonstrate a reduction in diversity of cellular clonal populations on PDX passage.

A similar loss in diversity in breast cancer PDX models was reported by *Nguyen et al.*, (197). Indeed, it was found that in one cell line xenograft model, the total number of detectable clones decreased from 185 to 12 in just three passages. Differences in growth rates and/or CIC between clones could preclude these models from use in experimental setups of a similar design to those employed in tamoxifen resistance studies detailed in *Chapter 3*.

Figure 5.3.1 – STG282CT model expansion

*Shows our STG282CT expansion and pooling strategy to increase barcode library representation. 900,000 individual lenti-ClonTracer PDTs at 64% viability by trypan blue exclusion (348) were subcutaneously injected into the flank of an NSG host, the resulting model - STG282CT-x0 – was engrafted into 10 NSGs (STG282CT-x1) before PDTs were harvested, mixed and re-implanted into 10 further NSGs (STG282CT-x2) in an effort to increase library representation. PDX models were expanded down specific lineages as detailed in the figure. Lines denote the flow of PDX material into subsequent hosts, red coloured models were analysed by NGS (for barcode composition) and red triangles represent suspected sites of metastasis. **Figure Overleaf.***

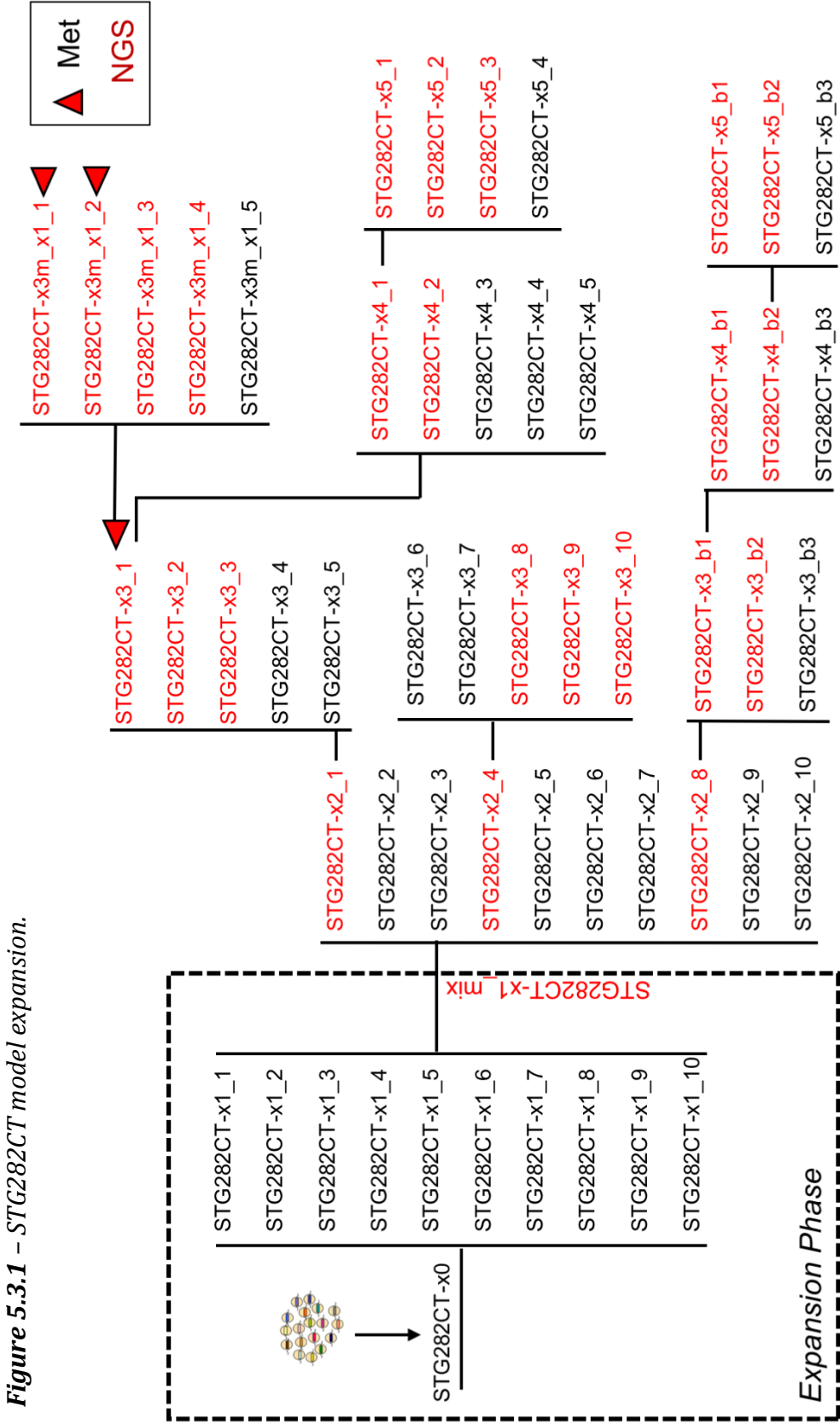


Figure 5.3.2 – Barcode Complexity in STG282CT-x1 pool

A) shows our STG282CT expansion and pooling strategy to increase barcode library representation. **B)** Despite a theoretical complexity of 900,000 uniquely marked cells, with up to 500-fold representation, only 1,034 unique barcode sequences were found in a sampling of mixed cells from 10 STG282CT-x1. In total, 1×10^6 cells from this pool were sequenced, to a depth of approximately 120×10^6 reads, with resultant analysis showing approximately 57×10^6 barcode sequences falling into 1,034 unique sequences. Second-generation tumours (STG282CTx2_1, STG282CTx2_4 & STG282CTx2_8) harboured a total of 1,231 unique barcode sequences, of which 578 were unique to passage x2 models and 653 were also seen in STG282x1. **C, D & E)** Show the relative abundance of the top 100 unique barcodes, i.e. the fraction of usable reads in each sample represented by each of 100 unique barcodes across the three samples. In each pairwise comparison, there is significant overlap in barcode abundance ($p < 0.001$), though the correlation is far from perfect ($r^2 = 0.35, 0.11$ & 0.41 for panels C, D & E respectively). **Figure Overleaf.**

Figure 5.3.2 – Barcode complexity in STG282CT-x1 pool

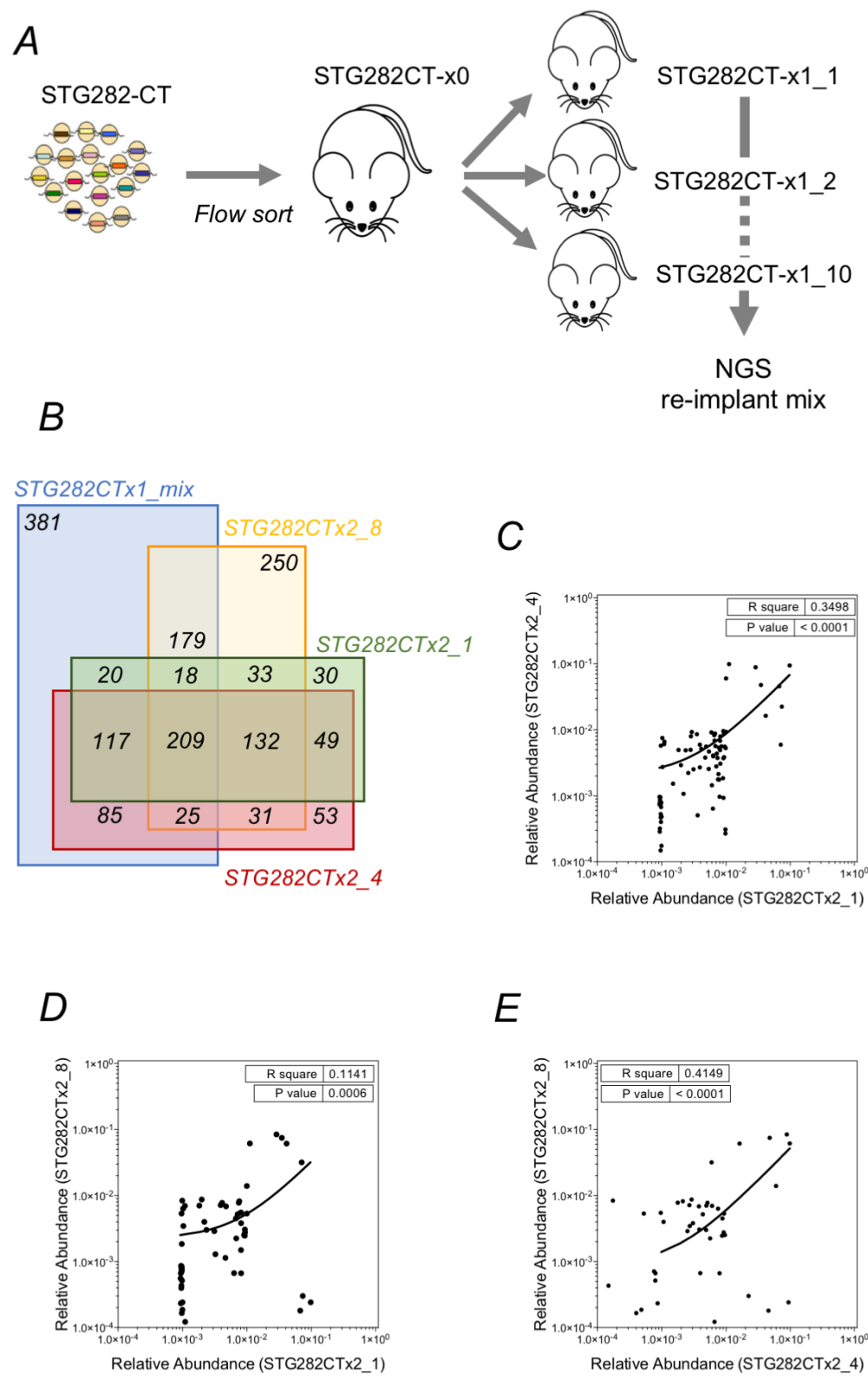


Figure 5.3.3 – Estimating cell number from barcode abundance

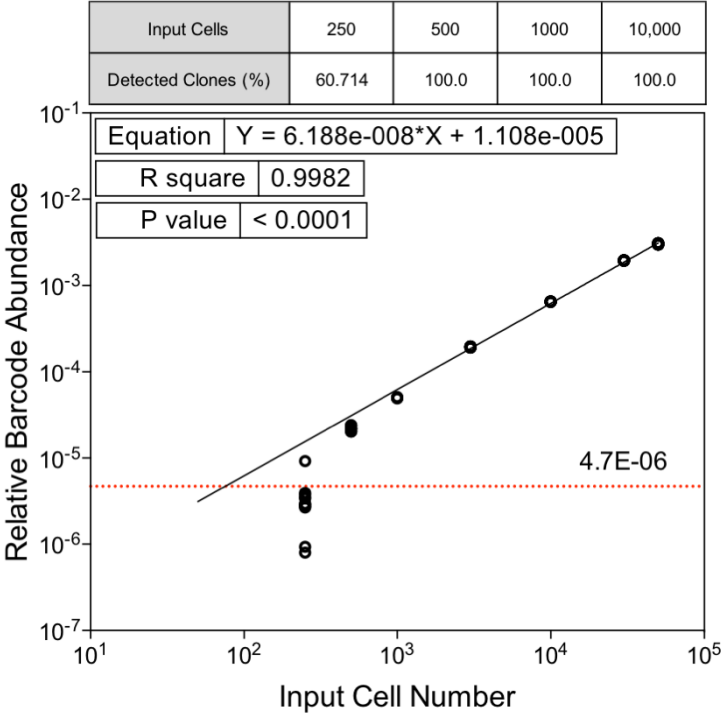
A) HEK293s infected with clonally derived (i.e. single barcode sequence) ClonTracer constructs were spiked into STG282CT samples prior to DNA extraction and NGS barcode sequencing. Across 29 sequenced PDX samples, barcodes spiked in at 500 or more cells were detectable in 100% of cases, and a good correlation was observed between the number of cells input and relative barcode abundance ($r^2 = 0.998$, $p < 0.0001$). Below 500 cells, the correlation between input cells and barcode abundance was markedly less clear, and our ability to detect clones fell to 60%. We set a threshold of 500 cells (roughly equivalent to an RBA of 4×10^{-6}), as our limit of confident detection in further studies. Importantly, we used spike in controls present in each sequenced sample to locally optimise this limit of detection and act as an internal control for barcode abundance experiments. The limit of detection (in terms of RBA value) is therefore slightly varied in different models. **B)** We used the correlation between input cell number and relative barcode abundance to calculate estimated cell number for spike in controls and found good concordance with the known number of spike-in cells, across a wide range of values (~1,000 to 50,000 cells; $r^2 = 0.998$, $p < 0.0001$), though the estimated clone size fell markedly below actual clone size with 500 or fewer spike-in cells. The relationship between relative barcode abundance and clone size across all samples is as follows:

$$\text{Clone Size} = (\text{RBA} - (1.108 \times 10^{-5})) / (6.188 \times 10^{-8})$$

Figure Overleaf.

Figure 5.3.3 – Estimating cell number from barcode abundance

A



B

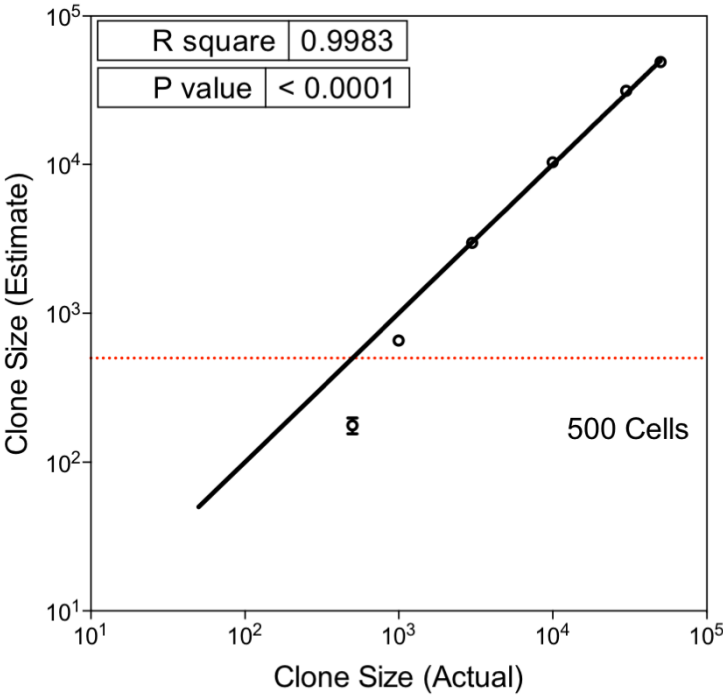
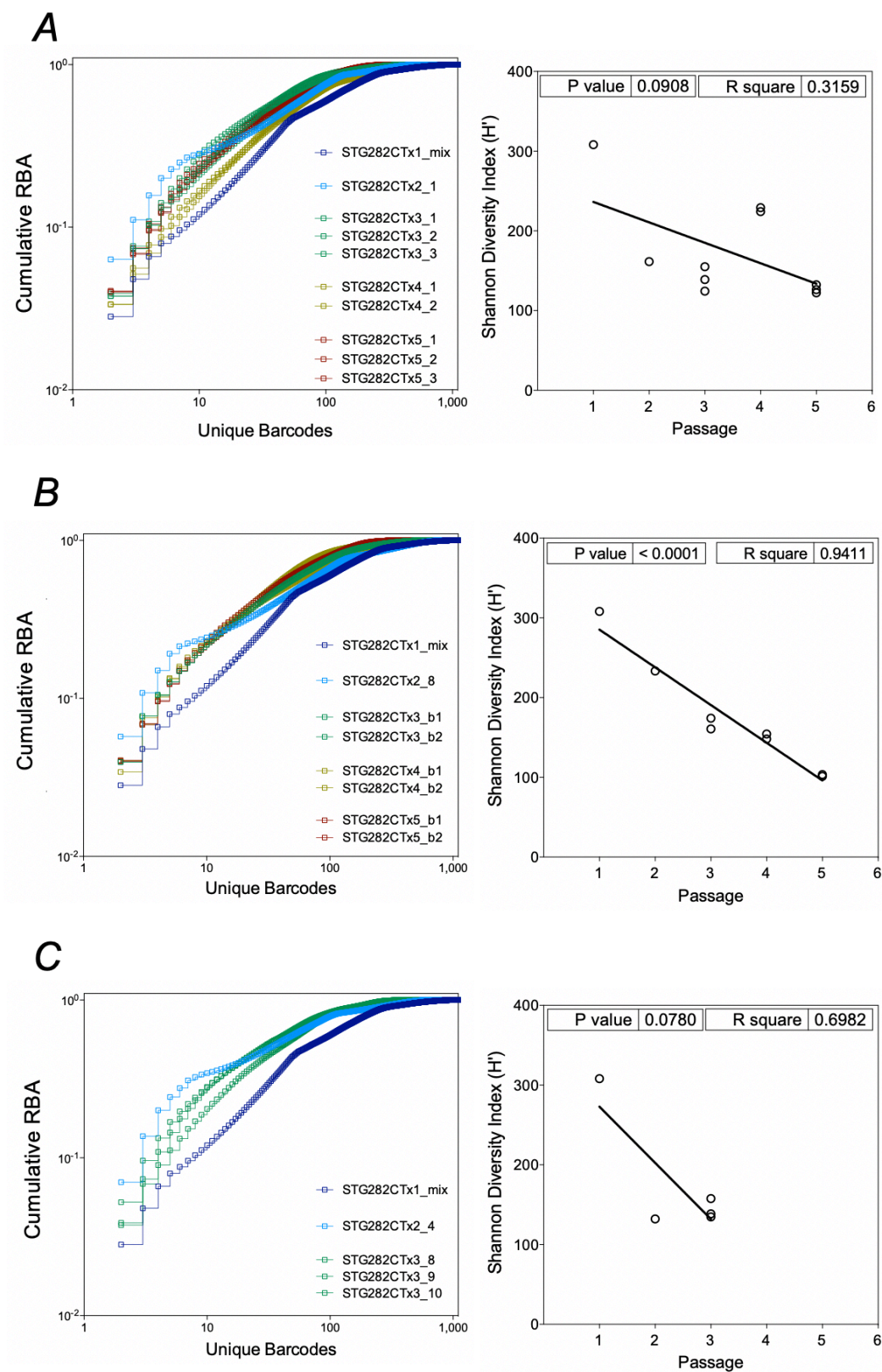


Figure 5.3.4 – Passage reduces clonal barcode complexity

*Cumulative relative barcode abundance (RBA) as a function of the number of unique barcodes counted, alongside the Shannon Diversity Index (SDI; H') (204,363). A, B & C) show three parallel passage arms from our PDX expansion map (Figure 5.3.1). **Figure Overleaf.***

Figure 5.3.4 – Passage reduces clonal barcode complexity



5.4 STG282-CT repopulation dynamics

PDX models have been shown to be remarkably stable over serial passage by ourselves (198) and others (195). However, certain changes in population growth and/or clonal dominance have been observed. In our own models, we have shown the majority of outright clonal selection occurs on initial implantation, but that approximately 20% of clonal populations, defined by PyClone (199), have significant changes in cellular prevalence across several years of *in vivo* culture and maintenance. Interestingly, we could attribute a putative driver mutation to only four of the 38 clonal clusters that changed significantly during serial passage (based on a ratiometric method to identify and classify driver events proposed by *Vogelstein et al.*, (452)). Moreover, in analysing matched PDX models derived from primary and metastasis, or from multiple punch biopsies of the same metastatic site, we have shown remarkable similarities between clonal clusters. In at least one case (STG139), shared clusters contained ~80% of all SNVs detected, comprising putative truncal clusters and sub-clonal clusters with estimated cellular prevalence <5% (198).

However, in our studies, and in those reported in the literature, changes in clonal prevalence across passages has been observed. One study sought to determine whether such directional clonal kinetics were associated with deterministic or stochastic processes, such as random genetic drift (195). The authors found that 80% of the directional changes seen in clonal structures were repeated across multiple parallel passages of the same PDX model. This included the reproducible expansion of initially minor subclones, implying a high likelihood of shared deterministic mechanism rather than repeated rare stochastic events (195). Conceptually, this could be linked to both acquisition and reduction of selective pressures on transplantation from human to murine host. For example, mutations aiding in the suppression of the immune response would likely no longer confer a fitness benefit in immunocompromised mice, meaning

that clonal populations able to suppress the immune system would no longer have a selective advantage and could therefore be outcompeted (23).

In cellular clonal tracing experiments with lentiviral barcodes, variable clonal dynamics have been associated with the emergence of resistance to chemotherapies. *Kreso et al.*, found several repopulation behaviours in CRC PDX models, certain cells reproducibly appearing in one passage before falling below the limits of detection in subsequent passages (196). In examining breast cancer PDX model repopulation dynamics in the unperturbed state, *Nguyen et al.*, (197) found both symmetric and asymmetric growth patterns in paired replicate models. Interestingly, the authors found xenografts derived from the MDA-MB-231 cell line demonstrated the most asymmetry in growth activity, with 33 to 40% of replicate clones showing different growth trajectories. Whereas in PDX models, 80 to 98% of replicate clones showed symmetry of growth activity, with the majority of asymmetry accounted for by clones decreasing in one passage-replicate and fluctuating in another (197).

We sought to understand the kinetics of clonal fluctuations in our models in more detail. *Figure 5.4.1* shows the trajectories of individual unique barcodes, depicted as grey lines connecting RBA at different passages. The 350 most abundant barcodes at each time point (representing >92% of cells) were plotted in this analysis for ease of interpretation. Dotted grey boxes connect matched implantations and black arrows denote the original sample from which the next passage (or matched passages/transplants) were derived. The limit of accurate detection (~500 cells) is shown in shaded grey at the base of each plot and in each panel, a PDX expansion map relating to the RBA plots, is depicted as an insert. In general, relatively flat lines denote stable clonal prevalence between two passages, downward trends represent clonal loss or reduction in clone size and upward trends represent resurgence of old clones, emergence of previously undetected clones, or clonal growth. Symmetric and asymmetric clonal growth is quantified for each set of matched implantations, in terms

of percentage of clones with neutral growth ('Neut'; RBA in reimplantation is within one SD of the mean fold change from originating sample), those that decrease in prevalence ('Down'; RBA is lower in reimplantation versus originating by >1 SD) and those that increase in prevalence ('Up'; RBA is higher in reimplantation versus originating by >1 SD).

Figure 5.4.1 A depicts such clonal dynamics for the STG282CT-x2_1 to STG282CT-x5 expansion arm. This is the largest collection of models studied together, both in terms of passage (covering four passages) and replicates, with both STG282CT-x2_1 and STG282CT-x4_1 being implanted into three replicates and STG282CT-x3_1 implanted into two. Interestingly, visual analysis shows similar repopulation dynamics across paired implantations. For example, STG282CT-x2_1 implantation reliably results in the loss of a group of clonal populations, the maintenance of a proportion of clones at around RBA of 10^{-3} , and the enrichment of few previously undetected barcodes. Implantation of STG282CT-x4_1 into three replicate models likewise almost universally results in the loss of (and /or downward trend of) the same clones, but a relatively stable overall trend. STG282CT-x3_1, shows a larger proportion of divergent clonal repopulation dynamics in the paired implantations. In one matched replicate (STG282CT-x4_1), a great deal of new clones were detected, and in the other (STG282CT-x4_2), the majority of clones were relatively stable, coupled with slight loss and/or downward trend of clones typically seen to be stable in prior generations. A smaller proportion of clones showed an upward trend in STG282CT-x4_2 than STG282CT-x4_1. These observations are quantified, in terms of the percentage of clones showing divergent (or asymmetrical) and symmetrical growth patterns (positive, neutral or negative) between replicates. In STG282CT-x3_1, all clone behaviours were largely symmetrical. STG282CT-x4_1 showed $>75\%$ of downward and upward trending clones following the same pattern, while those with a neutral path were 56% likely to follow this path across all three replicates, similarly, STG282CT-x2_1 showed an overall majority of symmetrical clonal behaviours.

In *Figure 5.4.1 B* the expansion of STG282CT-x2_8 is followed through two passages. Generally, clones present in STG282CT-x3_b1 follow similar patterns of symmetric growth on transplantation, though approximately 70% of clones with an upward trend are asymmetric. This can largely be attributed to the detection of 78 new clones in one model but not the other. Reimplantation of STG282CT-x4_b2 results in a majority of asymmetric growth patterns across all three type of clonal growth behaviours. Finally, *Figure 5.3.1 C* shows this set of analysis for the STG282CT-x2_4 expansion arm. In this case, the majority of cellular populations were neutrally trending on passage. Symmetrical growth patterns between the three analysed replicate models were seen in 92% of neutrally trending populations, 65% negatively trending and 83% positively trending. Interestingly, barcodes in the first passage of this arm appeared particularly tightly grouped, with the loss of the majority of barcodes with an RBA of $\sim 7 \times 10^{-3}$ and the emergence of a tight cluster of barcodes all growing to an abundance of $\sim 9 \times 10^{-4}$. Overall, our barcoded PDX models showed both asymmetrical and symmetrical growth patterns between replicates established from the same originating tumour. Growth patterns themselves were varied across populations and between independent expansion of STG282CT. In later passages, we observed repeated detection of barcodes not seen in earlier models, together with consummatory loss of populations which had been relatively stable across multiple passages.

Figure 5.4.1 – STG282CT Repopulation dynamics

*Depicts the trajectories of individual unique barcodes, grey lines connecting their RBA at different passages. The 350 most abundant barcodes in each passage (representing >92% of cells) are plotted for ease of interpretation. Dotted grey boxes connect matched or repeat passages and black arrows denote the original sample from which replicate passages were derived. For each parallel engraftment, we compare symmetric and asymmetric repopulation dynamics. 'Neut'; RBA in reimplantation is within one SD of the mean fold change from originating sample, 'Down'; RBA is lower in reimplantation versus originating by >1 SD, and 'Up'; RBA is higher in reimplantation versus originating by >1 SD. The limit of accurate detection (~500 cells) is shown in shaded grey at the base of each plot and in each panel a PDX expansion map relating to the RBA plots shown, is depicted as an insert. Importantly, we used spike in controls present in each sequenced sample to locally optimise this limit of detection and act as an internal control for barcode abundance experiments. The limit of detection therefore slightly varied in different models. **A, B & C**) Show three parallel passage arms from our PDX expansion map (Figure 5.3.1). **Figures Overleaf.***

Figure 5.4.1 A – STG282CT Repopulation dynamics

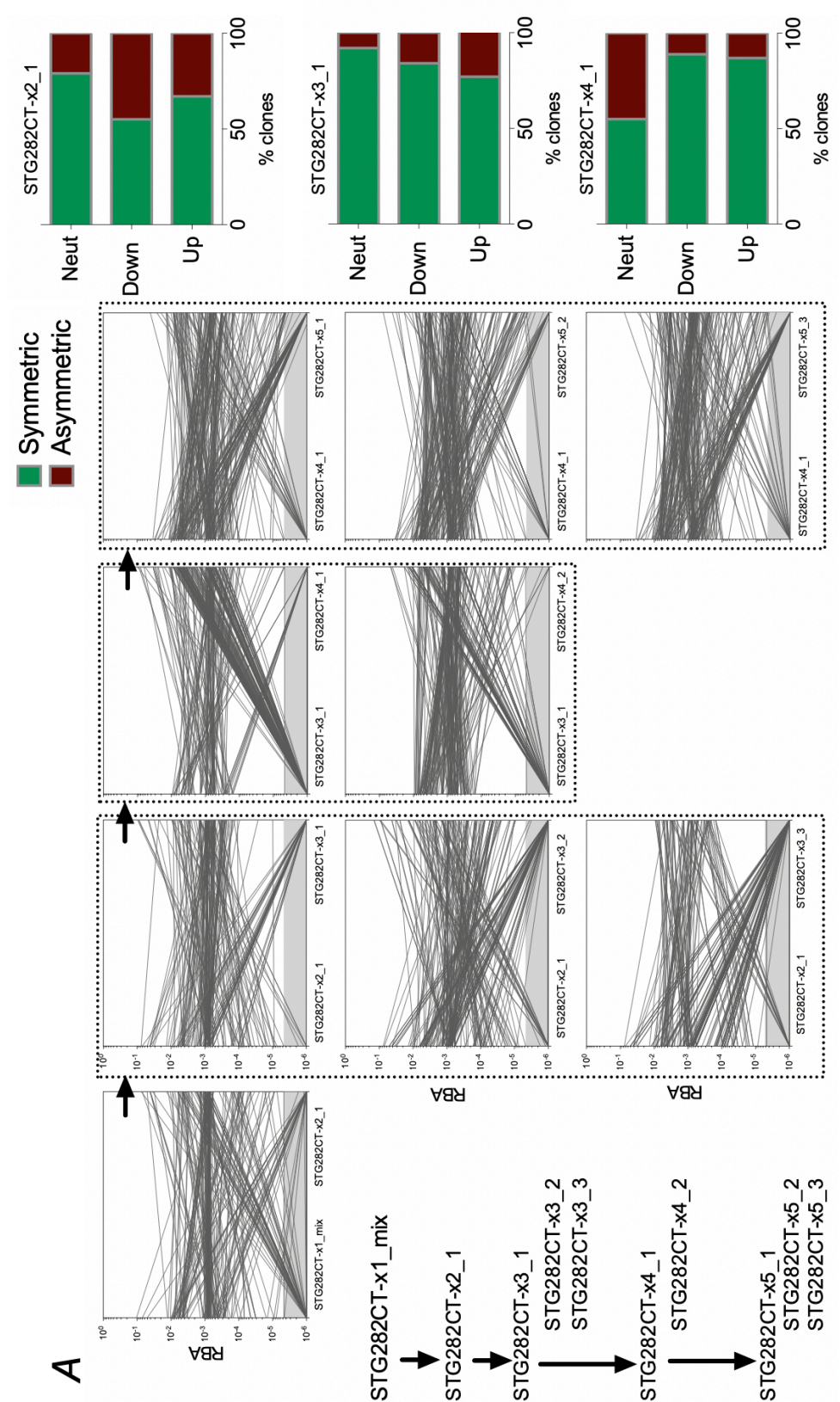


Figure 5.4.1 B – STG282CT Repopulation dynamics

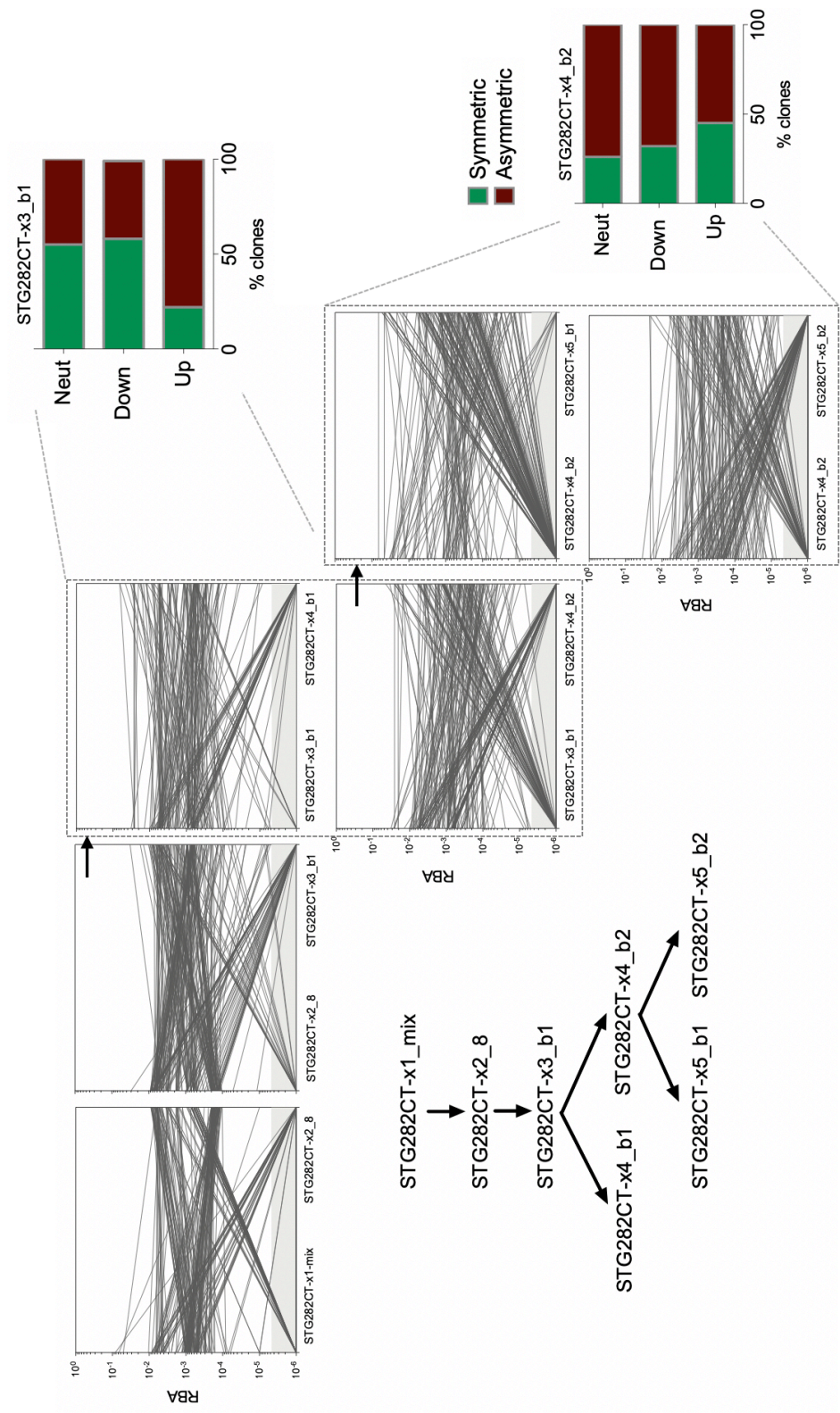
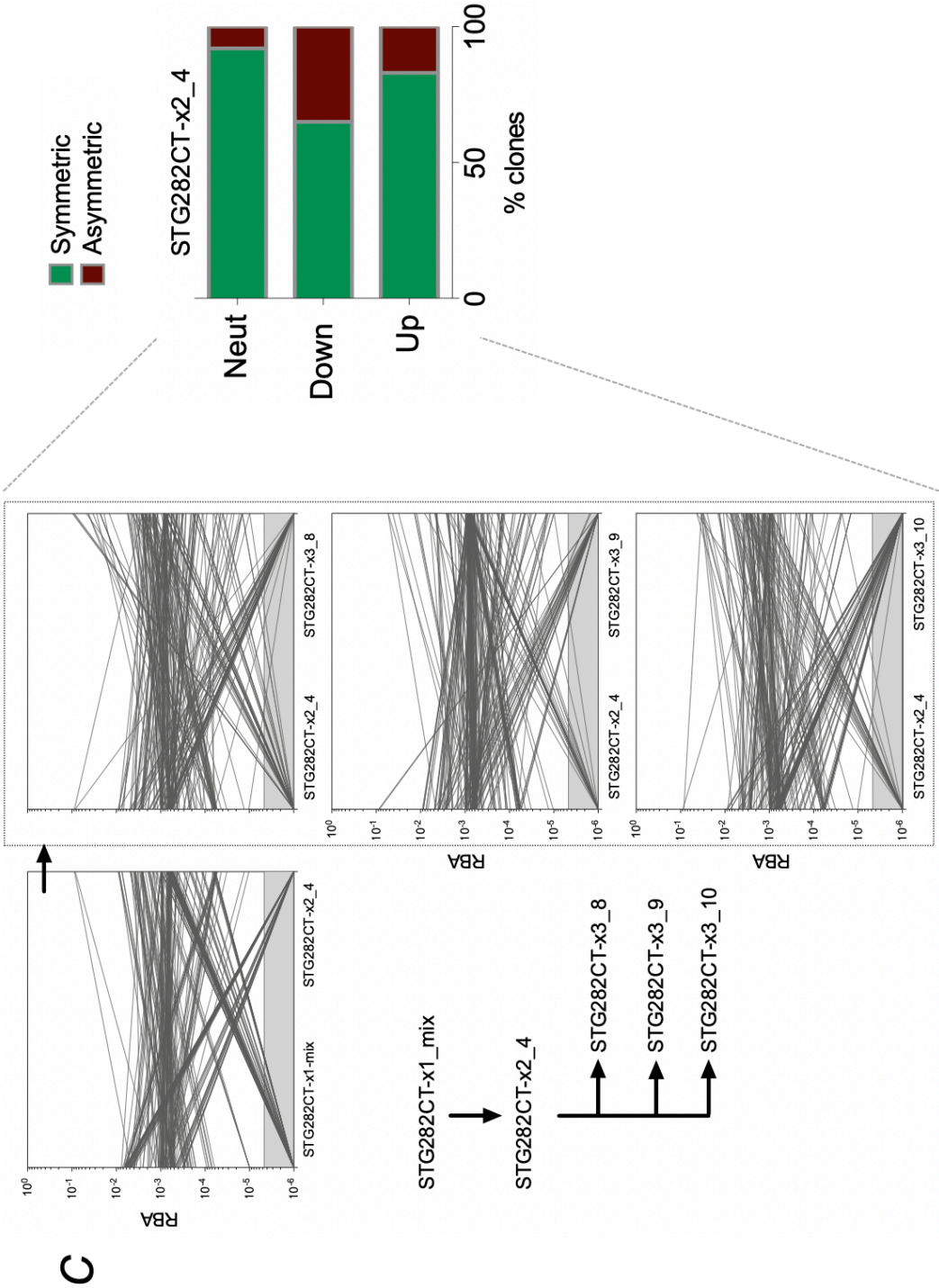


Figure 5.4.1 C – STG282CT Repopulation dynamics



5.5 Phenotypic subtyping of barcoded clones

In *Figure 5.4.1* we observed multiple distinct clones that were initially below our limit of detection (~ 500 cells) reproducibly appeared in multiple replicates of later passages. In some cases, these grew to $>1.6 \times 10^5$ cells in a single passage, suggesting that we were not simply passaging slow growing clones or non-viable cells between models, but were witnessing divergent growth kinetics of the same cells over time. This prompted us to further investigate the presence of distinct clonal phenotypes present in our models, based on their repopulation kinetics. Such subtypes have been described previously by multiple accounts in the literature.

For example, *Kreso et al.*, defined five lentiviral clone types in a panel of CRC PDX models, with an experimental procedure relying on lentiviral insertion site mapping. Their experimental design had an estimated limit of detection (by Southern blot) of 10^4 cells (196). *Type I* (long-term persistent clones) were defined as those detected in reliably in each passage. *Type II* were defined by their presence in first and second passages, but absence in the last passage, demonstrating clones with limited CIC. *Type III*, transient clones, were defined by detection in the first passage, but not again in any later passage, irrespective of how many additional transplants were carried out. *Type IV*, quiescent or resting clones, were defined by their absence from early passages, but presence in at least one later passage. These clones were characterised by slow, sustained growth, and resistance to chemotherapy. Fluctuating, or transient growth, behaviour was attributed to *Type V* clones, which were detected in one passage, absent from another, and detected again in a following passage (196). *Nguyen et al.*, likewise defined several clusters based on the kinetics of repopulation in their breast cancer cell-line xenografts and PDX modes (197). In this case the authors relied on k-means clustering, a technique where n observations are partitioned into k clusters so as to minimize within-cluster variance (453). For each experiment, the authors thus defined several distinct kinetic behaviours. For MDA-MB-231 derived xenografts, three clusters were defined: *i*) a relatively stable

clone size, *ii*) a relative clone size increasing on passage and *iii*) a relative clone size decreasing in the second passage. In a second MDA-MB-231 experiment, a further two clusters were defined, both characterized by delayed clonal growth. One cluster was further characterised by being detectable in secondary passages but decreasing in tertiary passages. SUM-149 derived xenografts, despite being expanded over a single passage *in vivo*, yielded six and seven clusters, though these were variations of the same neutral, positive and negative growth trends. PDX models analysed by the group shared only three of the patterns exhibited by the serially passaged cell lines; no change, decreasing or fluctuating; but not clones that increased in size or showed delayed growth onset (197).

We noted that the presence of clusters beyond the five identified by *Kreso et al.*, was perhaps not additive to our biological understanding of PDX repopulation kinetics or clonal fitness. Furthermore, discrepancies between these and those described by *Nguyen et al.*, could be partially attributed to experimental design. Barcoded MDA-MB-231s were expanded, at most, into two subsequent passages, SUM-149 into one subsequent passage, and patient derived models (T1, T2 & T3) into two, one or zero subsequent passages (197). This relatively short follow-up makes detection of *Type IV* and *V* clones particularly challenging. Moreover, *k*-means clustering is heuristic, meaning it is best suited for clustering Gaussian distributions, and may not be as accurate when grouping cellular growth patterns, which are typically exponential in nature (454). In our experiments, we opted to define clones based on the *Kreso et al.*, classification. However, 86 unique barcodes failed to meet the criteria for classification in any of the five clusters. This tended to be attributed to limitations in the experimental design: for example, unique barcodes falling in the STG282CT-x2_4 expansion arm, with loss of detection in STG282CT-x2_4, could be attributed to *Type III* had a fourth passage been included in this arm. Going forward, we could develop a bespoke model-based clustering approach to remove any potential sources of bias and fully cluster the entirety of our unique barcoded clones.

The results of our clonal phenotype analysis are shown in *Figure 5.5.1*. Individual unique barcodes seen in any passage or replicate of STG282CT expansion experiments were classified into either *Type I, II, III, IV* or *V* based on repopulation kinetics. *Figure 5.5.1 A* shows *Type I* clones, *i.e.* those detected in each passage of each arm. Contained in the colour coded boundary box, a schematic of *Type I* inclusion criteria (*top*), RBA across the three independent STG282CT expansion arms (*left*) and the RBA of each barcode shared between expansion arms, expressed as the median across repeat transplantations and parallel expansion arms (*right*) is shown.

In total there were 824 unique *Type I* clones defined, ranging in cell content from 628 to 5.38×10^6 cells across two sequential samplings (passages). Though prevalence of *Type I* clones as a whole was relatively stable across multiple passages, it was not uncommon for fold changes in the order of 1-2 orders of magnitude (in either direction) when we consider individual clones. *Figure 5.5.1 B* shows *Type II* clones, displayed in a similar fashion to *Type I* clones in *Figure 5.5.1 A*. *Type II* clones were defined as those detectable across each expansion arm in x1, x2 and x3 model passages, but were not detected in any passage 5 (x5) arm or replicate. In total, 524 clones matched these criteria in STG282CT-x2_4 and STG282CT-x2_8 expansions, with 392 also present in the STG282CT-x_4 expansion arm. As this arm did not progress to passage 5, we could not be sure that they would also show *Type II* behaviour in this arm. *Figure 5.5.1 C* shows the same schematic for identification of 432 early not-detected *Type III* clones, each present in STG282CTx1 and STG282CTx2, but not in passage x3 or any subsequent passage. The majority of these *Type III* clones only showed a net loss (*i.e.* median RBA $< \sim 4 \times 10^{-6}$) from passage x3 onward, whereas a great deal of heterogeneity across passage x2 replicates was seen, with 32%, 17% & 48% of sequences detected in STG282CTx1 lost in STG282CT-x2_4, STG282CT-x2_1 & STG282CT-x2_8, respectively. *Figure 5.5.1 D* shows *Type IV* clones, defined by an RBA below the limit of detection until at least STG282CT-x3, at which point they remained detectable in each subsequent passage. 312 unique *Type IV* clones were detected across our profiled models. Finally, *Figure 5.5.1 E* shows *Type V* clones, defined

by fluctuating repopulation kinetics. In this case, barcodes which followed the pattern or either i) *detection, loss, detection, loss* ii) *detection, loss, loss, detection* or iii) *loss, detection, loss, loss* in any sequence of four or more passages, were defined as *Type V*, and 459 total barcodes fit these criteria.

Having associated each cellular clone with a specific pattern of repopulation kinetics, we next sought to investigate their relative contributions to the cellular population in each passage of our barcoded PDX model. *Figure 5.5.2* shows the percentage of each clonal phenotype (*Type I, II, III, IV & V*; defined in *Figure 5.5.1*) by total cell number. Inserts show the total number of unique barcodes seen in the respective sample, while arrows denote the path of transplantation and boxes separate replicate implantations. *Figure 5.5.2 A* presents corresponding data for the STG282CT-x2_1 expansion arm, *Figure 5.5.2 B* for STG282CT-x2_8 and *Figure 5.5.2 C* for STG282CT-x2_4. Not surprisingly, in *Figure 5.5.2 A*, we see a progressive loss of first *Type III*, then *Type II* clones in passage x3 and x4 respectively. Loss of *Type III* clones is accompanied by a reduction in the number of unique barcodes across STG282CT-x3_1, STG282CT-x3_2 & xSTG282CT-x3_3, from 608 to 444, 385 & 383, respectively. However, the loss of *Type II* clones on the following transplantation (*i.e.* in STG282CT-x4_1 & STG282CT-x4_2) is countered by substantial expansion of *Type IV*, slow growing or quiescent, clones and an accompanying increase in overall barcode complexity in these samples (to 659 & 650). Surprisingly, the fifth passage in this arm reverses this pattern again. *Type I* clones are reduced in overall cellular representation, whereas *Type V* transient clones increase in proportion, and *Type IV* clones remain relatively stable. Overall clonal count drops from 659 in STG282CT-x4_1 to 380, 327 & 305 in STG282CT-x5_1, STG282CT-x5_2 & STG282CT-x5_3, respectively.

Thus, there is a reduction in total number of clones, representation of *Type I* clones (typically stable) and an increase in *Type V* clones. This trend can be rationalised by analysing the estimated cell count of each population: STG282CT-x4_1 has 102 unique

barcodes belonging to *Type I* clones, 58 *Type IV* clones & 497 *Type V* clones. On passage, there was a ~2.5-fold reduction in the absolute number of *Type V* barcodes (to 225, 164 & 139), with the remaining *Type V* clones showing a near ~3-fold increase in cell content (from an average of 4,500 cells per barcode to 13,000, 15,550 & 15,200). This was accompanied by a reduction in *Type I* complexity (modestly, from 102 to 91, 92 and 97 barcodes) and average size of *Type I* clones (from 18,500 to 11,500, 8,150 & 7,900). This shift in overall population dynamics between passage x4 and x5 resulted in the changes shown in *Figure 5.5.2 A*. *Figure 5.5.2 B* shows similar data but presented for the STG282x-2_8 expansion arm. We see the same overall pattern as in *Figure 5.5.2 A*, with a tendency toward reduced barcode complexity and increased dominance of *Type IV* and *V* clones on serial re-transplantation. This could simply be a feature of the clonal phenotype definitions; with *Type IV* & *V* obligated to occur in later passages and *Type II* & *III* to fall below the limit of detection in early passages.

In *Figure 5.4.1 B*, we found relatively discordant barcode dynamics between paired implantations of STG282CT-x3_b1 (asymmetric growth in 78% growing, 41% shrinking & 45% neutral). Similarly, we found poor concordance in STG282CT-x4_b2 (asymmetric growth in 55% growing, 68% shrinking & 74% neutral). In this secondary analysis, we now see the five clonal phenotypes are variably dominant in repeat implantations of STG282CTx3_b1 (*i.e.* STG282CT-x4_b1 & STG282CT-x4_b2). Indeed, the second most dominant clonal *Type* by cell number, representing ~3 x10⁶ cells, is *Type II* (29%) in STG282CT-x4_b1 and *Type IV* (34%) in STG282CT-x4_b2. In STG282CT-x4_b1 tumours, *Type IV* clones make up just 2% cells, and in STG282CT-x4_b2 tumours, *Type II* clones make up just 6%. We also found low concordance in growth patterns between replicate implants of STG282CT-x4_b2 in *Figure 5.4.1 B*, and in this case, we can attribute differences to an increase in *Type IV* clonal prevalence, and a reduction in the presence of *Type Vs* (*Figure 5.5.2 B*). Lastly, we applied this analysis to progeny of STG282CT-x2_4. *Figure 5.5.2 C* shows a similar pattern of reduction in detection of *Type II* and *Type III* clones, concurrent with a reduction in barcode complexity. The distribution of clonal

Types was remarkably similar across repeat implantations of STG282CT-x2_4, down to mirrored percentages of *Type IV* clones beginning to appear in passage x3 tumours (at 3%, 2% & 3%).

Overall, the analyses presented in this section have confirmed our earlier observation that matched transplants resulted in relatively similar repopulation kinetics in daughter tumours. By defining clonal populations based on the pattern by which they were observed to repopulate a tumour, we were able to better understand instances where matched transplants were discordant. For example, in asymmetric growth in STG282CT-x3_b1 matched transplants (into STG282CT-x4_b1 and STG282CT-x4_b2; *Figure 5.4.1B* & *Figure 5.5.2B*) was due to the appearance of a group of clones not previously seen in any model (*Type IVs*). Moreover, asymmetric growth in further reimplantation of STG282CT-x4_b2 was attributed to the loss of *Type V* clones in only one model, rather than a repetition of the *Type IV* clone selection seen in the previous generation. However, this methodology is limited by experimental design: a *Type IV* clone could be classified as a *Type V* had the experiment consisted of a further passage (presuming it was undetected), likewise, a *Type I* or *II* clone could in fact be a *Type IV*.

Figure 5.5.1 – STG282CT clonal phenotypes

*Shows individual unique barcodes seen in any passage or replicate of our STG282CT expansion experiments, classified into Type I, II, III, IV & V clonal behaviours. Putative clusters had to fulfil the requirements of their clonal group in each of the three expansion arms, if present and if experimental conditions allowed, in order to be designated as a member of that clone. **A)** Shows Type I clones, i.e. those detected in each passage of each arm. Contained in the colour coded boundary box, Panel A shows a schematic of Type I inclusion criteria (top) and median RBA across the three independent STG282CT expansion arms (bottom). **B)** Shows Type II clones, displayed in a similar fashion to Type I clones in A. **C)** Shows the schematic for identification of 432 early not-detected Type III clones, each present in STG282CTx1 and STG282CTx2 in passage x1 or x2, but not seen in x3 or any subsequent passage. **D)** Shows Type IV clones defined by barcodes which were below the limit of detection until at least STG282CT-x3, at which point they remained detectable in each subsequent passage. 312 unique Type IV clones were detected across our profiled models. **E)** Shows Type V clones, defined by fluctuating repopulation kinetics. In this case, barcodes which followed the pattern or either i) detection, loss, detection, loss ii) detection, loss, loss, detection or iii) loss, detection, loss, loss in any sequence of four or more passages, were defined as Type V. 459 barcodes were included in this classification, with the remaining 86 unique barcodes failing to meet classification criteria for any of the five clusters. **Figure Overleaf.***

Figure 5.5.1 A, B & C – STG282CT clonal phenotypes

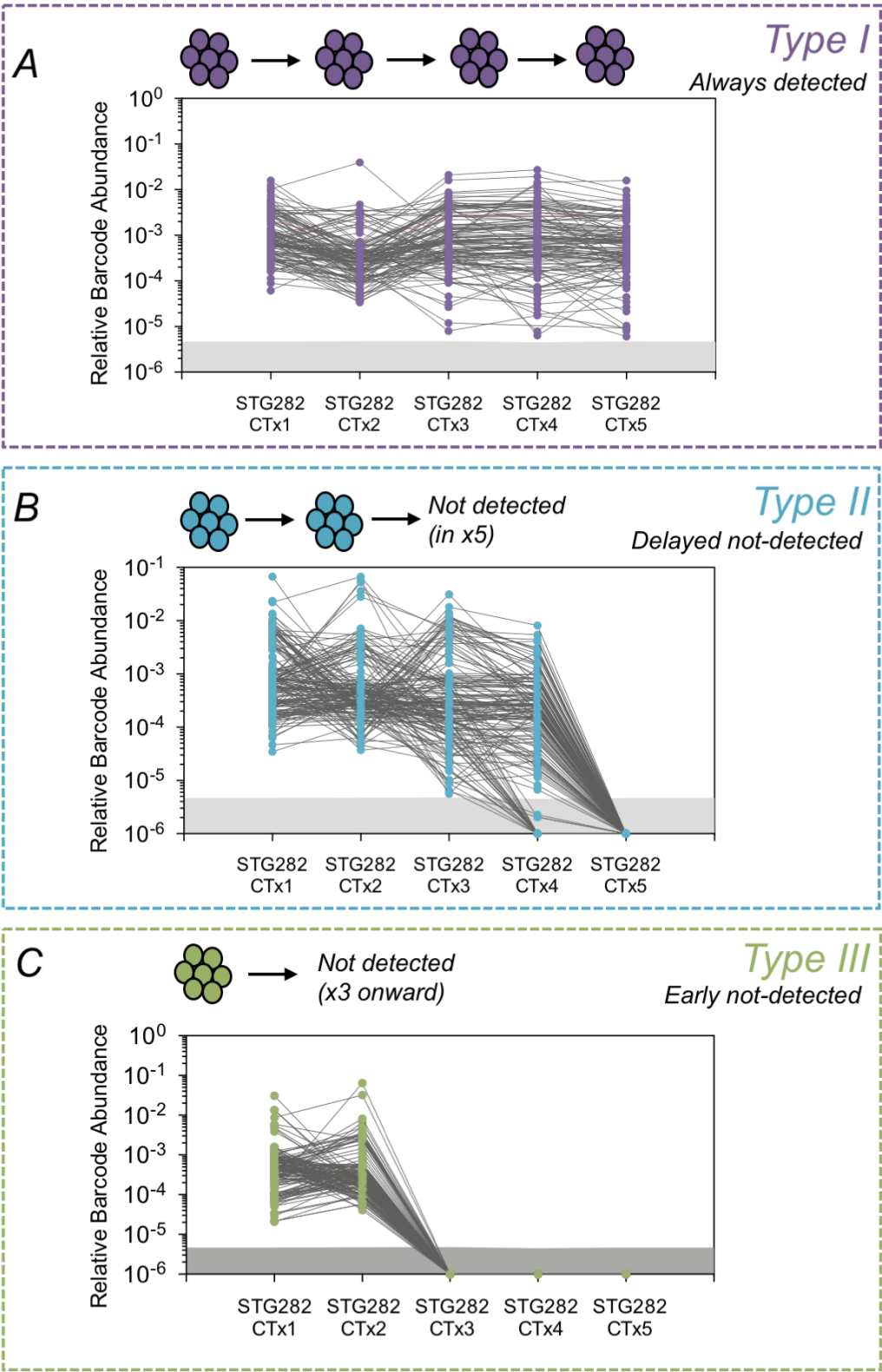


Figure 5.5.1 D & E – STG282CT clonal phenotypes

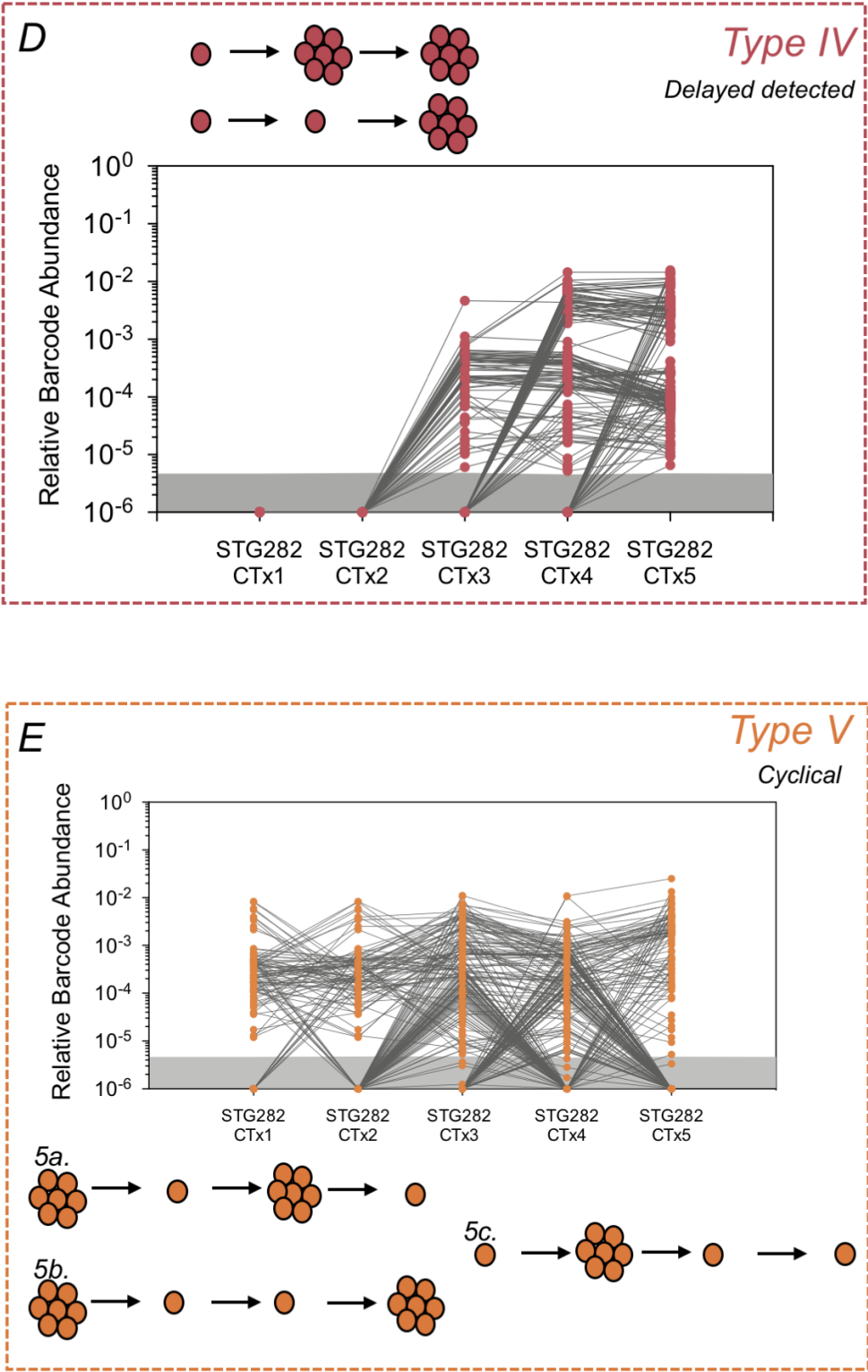


Figure 5.5.2 – Contribution of clonal phenotypes to repopulation kinetics

Shows the percentage of each clonal phenotype (Type I, II, III, IV & V; defined in Figure 5.5.1) by cell number. Inserts show the total number of unique barcodes seen in the respective sample, while arrows denote the path of transplantation and boxes separate replicate implantations. **A)** Presents corresponding data for the STG282CT-x2_1 expansion arm, **B)** For STG282CT-x2_8 and **C)** For STG282CT-x2_4. **Figure Overleaf.**

Figure 5.5.2 A – Contribution of clonal phenotypes to repopulation kinetics

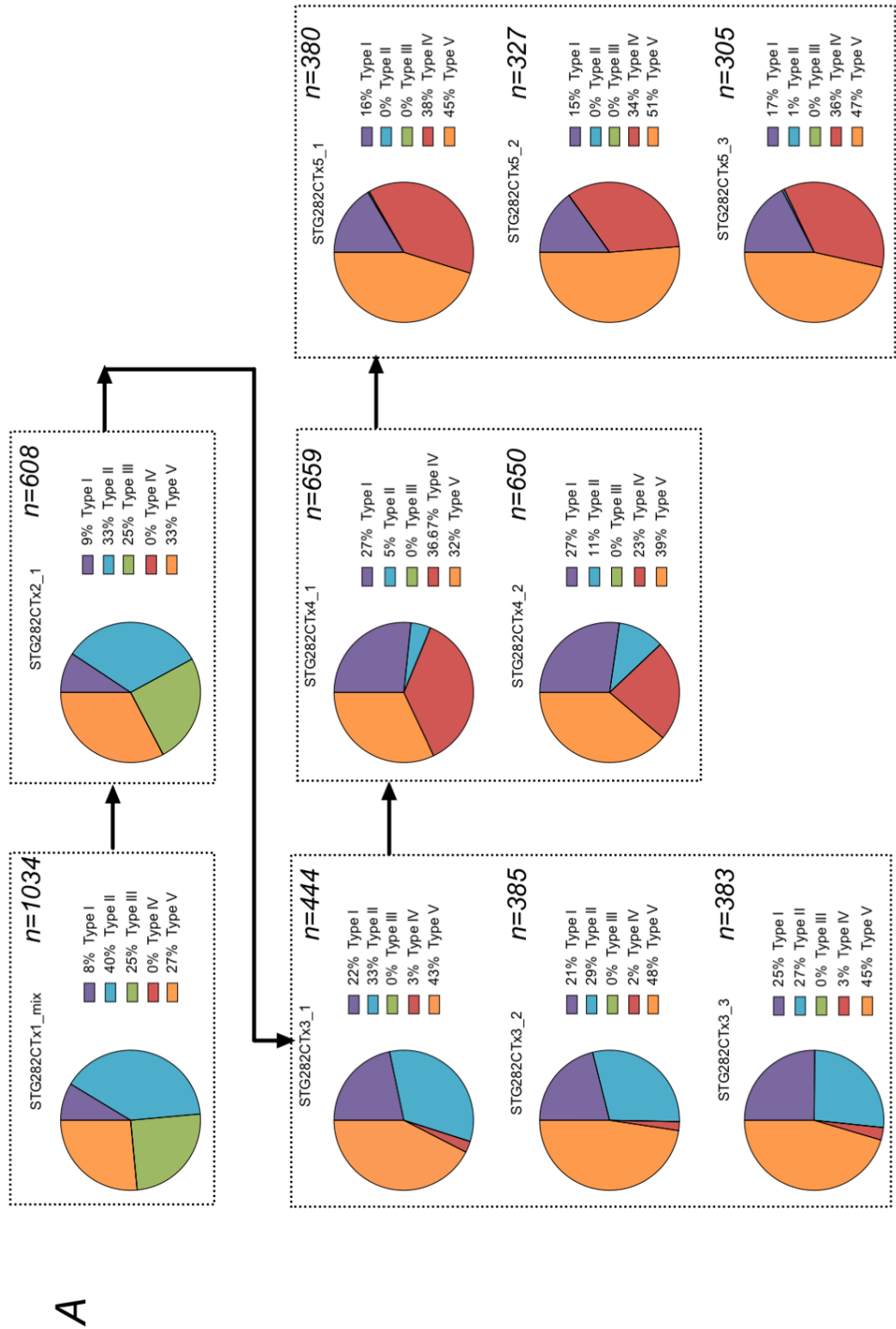


Figure 5.5.2 B – Contribution of clonal phenotypes to repopulation kinetics

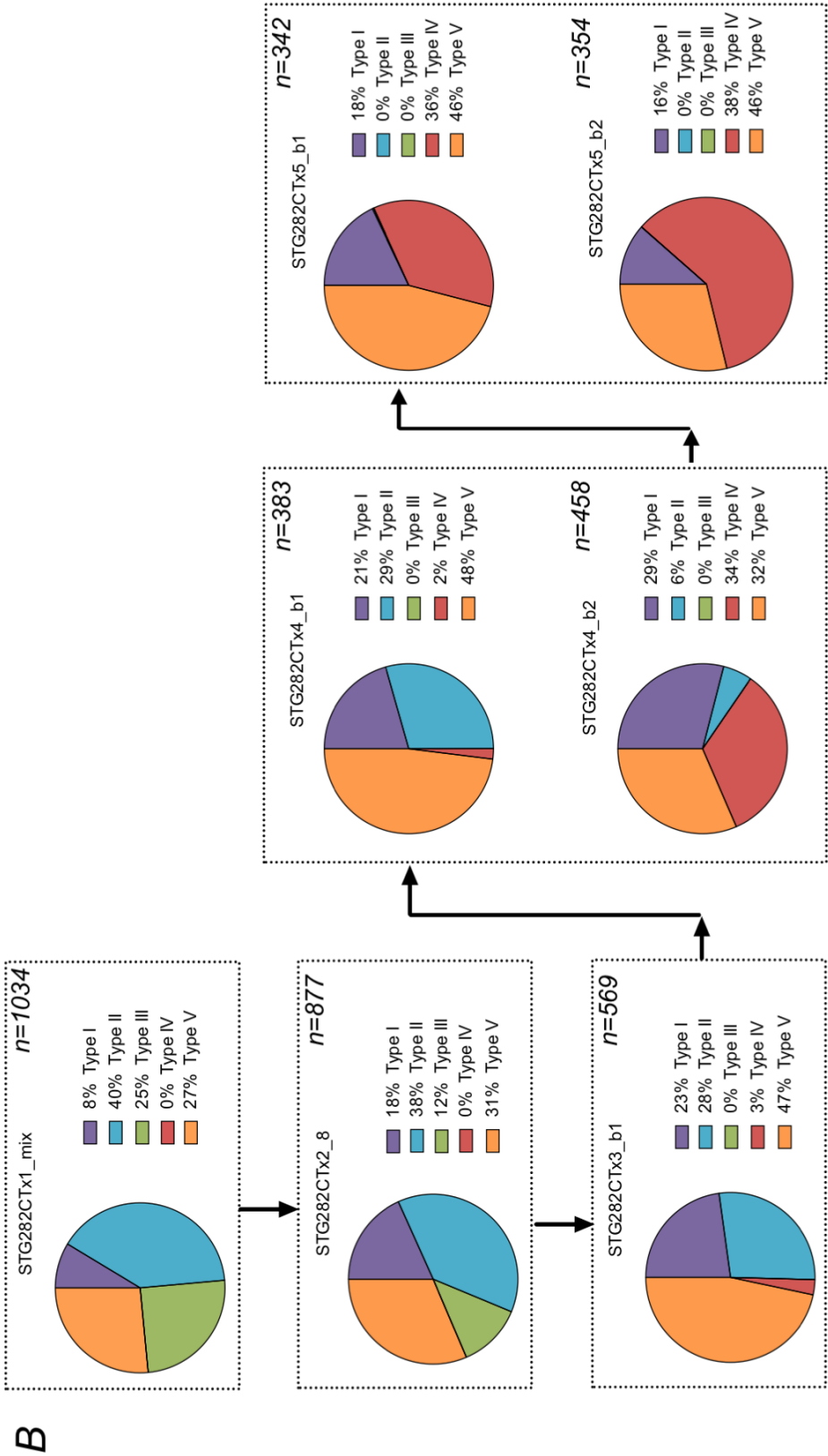
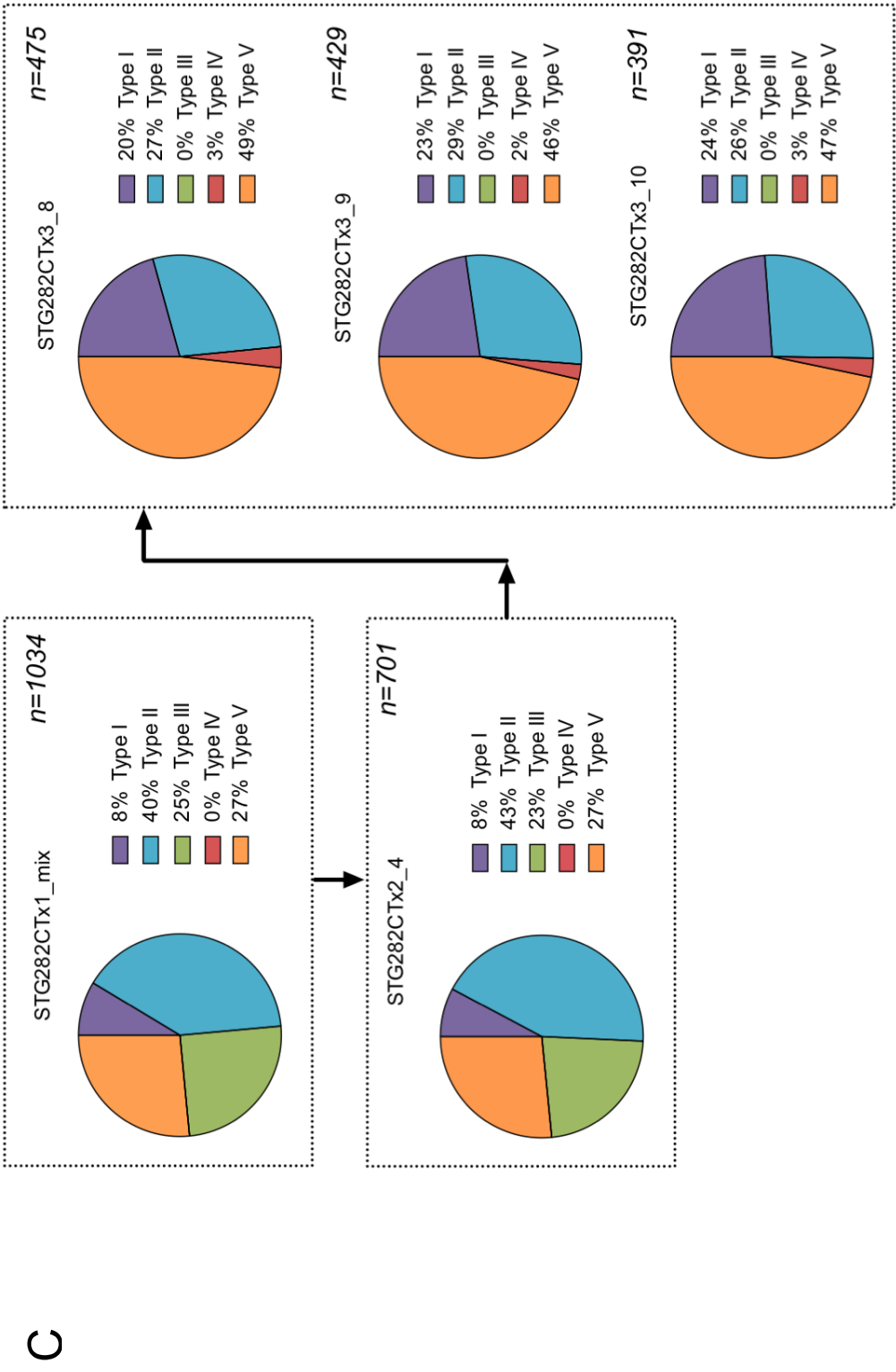


Figure 5.5.2 C – Contribution of clonal phenotypes to repopulation kinetics



5.6 Profiling of metastatic sites in STG282CT

STG282 was derived from the metastatic deposit of a TNBC. TNBC is a heterogeneous group of diseases, characterised and defined by the absence of ER α , PR and HER2 surface expression. Patients with TNBC typically have a relatively worse disease-free survival than those with other breast cancer subtypes, largely owing to an inherently aggressive clinical behaviour and a lack of recognized molecular targets for therapy (455). On average, TNBCs carry 1.68 somatic mutations per Mb of coding regions (~60 somatic mutations in each tumour). The mutation burden is not uniform across TNBC, and some tumours have a high mutation burden (more than 4.68 somatic mutations per Mb) (38,336), and a frequent occurrence of multiple copy-number aberrations involving genes that lead to multiple pathway alterations (7,38,73,78,198). Our group has helped delineate the diverse, variable genomic clonal composition of TNBC (456). By allelic frequency measurements of 2,414 somatic mutations in 104 TNBCs, a complete spectrum of molecular and clonal compositions was characterised at diagnosis (456). Aside from increased intratumour heterogeneity, and a lack of effective molecular targeted therapies, TNBC has been linked to an increased propensity for metastasis and the process of EMT (149,159).

Metastases is the ultimate cause of 90% of all cancer deaths (147), though questions remain around the origins and processes associated with metastatic spread. Distant metastases has been reported a late event in molecular time, though we reasoned that through lentiviral lineage tracing in our PDX models, a comprehensive study of breast cancer metastasis, together with annotation of metastatic-forming clonal populations could bring considerable advances to the field. Thus, we sought to identify potential metastatic deposits across our STG282CT expansion arms.

For each animal culled we performed a *post-mortem* procedure, detailed in *Methods, Section 2.1.5*. Briefly, PDTCs were seeded in the flank of NSG mice, leading to tumour

formation around the 4th and 5th mammary fat pads. Rapid dissection of the tumour area allowed samples from abdominal and inguinal mammary glands, together with suspected tumour draining lymph nodes (inguinal and/or lumbar) to be taken along with tumour harvesting. Lungs, liver and spleen were inspected (visually or with the Xenogen IVIS small animal imaging system) for potential metastatic sites. Harvested tissues could then be dissected and interrogated by IHC/histology if required. *Figure 5.6.1 A* shows a representative image of suspected micro-metastatic sites identified around the base of murine right lower lung lobule (*black arrows*), as imaged by the Xenogen system. These regions were dissected and stained with H&E for visual inspection. In each case where potential micro-metastatic sites were identified; H&E showed no obvious tumour structures. In total, we isolated and inspected >100 mammary glands and tumour draining lymph nodes before finding evidence of metastasis.

In one case (STG282CT-x3_1; *Figure 5.6.1B*) we identified a small cell mass (*arrow 2*) at the base of the 4th (abdominal; *arrow 3*) mammary gland near the site of tumour fragment implantation (*arrow 1*). Though this mass was not physically attached to the tumour bulk, it could have resulted from cells becoming dislodged during implantation and may not in fact be a metastasis. Nevertheless, the fragment was isolated and dissociated into six equally sized fragments, one for NGS and barcode analysis, five for reimplantation into NSG mice. *Figure 5.6.2* shows a schematic for expansion of this local metastasis (designated STG282CT-x3m) together with the results of ClonTracer barcode analysis. Barcodes seen in previous samples (98% of those identified) were assigned a clonal *Type (I-V)* based on repopulation dynamics in *Figure 5.5.1*. Interestingly, substantial differences existed between the proportion of clonal phenotypes in STG282CT-x3_1 and its matched metastatic sample, STG282CT-x3m. *Figure 5.6.2* shows STG282CT-x3m contains fewer total barcodes than STG282CT-x3_1 (345 versus 444), and that proportions of each clonal phenotype are markedly different. STG282CT-x3_1 is primarily comprised of *Type II* (33%) and *Type V* (43%) clones, with a sizable

proportion of stable *Type I* clones (22%), quiescent *Type IV* clones make up only 3% of the total cell number. Whereas, STG282CT-x3m is dominated by *Type IV* clones (77%). When STG282CT-x3m was re-implanted into STG282CT-x3m_x1 mice, we see a reduction in the proportion of *Type IV* & *V* clones, and an increase in *Type I* clones.

Of note, two out of five STG282CT-x3m progeny formed similar suspected metastasis in local mammary glands (as in *Figure 5.6.1*). Prior to this point, STG282CT-x3m was the first recorded metastatic lesion found in our analysis of >100 glands and >50 total instances of STG282CT injection into NSG mice. Hence, we reasoned that serial implantation of STG282CT-x3m was enriching for a metastatic phenotype. Interestingly, 2nd generation metastatic nodes (designated STG282CT-x3m_x1_1m & STG282CT-x3m_x1_2m) comprised a similar proportion of *Type IV* clones to STG282CT-x3m (79% & 85% respectively).

To quantify any specific enrichment within phenotypic clonal *Types I-V*, we next compared the relative abundance of 152 shared barcodes in paired metastatic and primary STG282CT models. *Figure 5.6.3 A* compares barcode abundance in secondary metastases to their paired non-metastasis primary tumour (STG282CTx3m_x1_1m in the left-hand panel; STG282CTx3m_x1_2m in the left-hand panel), poor correlations are seen in each case ($r^2=0.0295$, $p=0.0876$ & $r^2=0.0075$, $p=0.39$). *Figure 5.6.3 B* directly compares abundance between secondary metastases, where a significant correlation is seen ($r^2=0.402$, $p < 0.0001$). *Figure 5.6.3 C* compares abundance in STG282CTx3m versus secondary metastatic sites (STG282CTx3m_x1_1m in the left-hand panel; STG282CTx3m_x1_2m in the right-hand panel), significant correlations are seen in each case ($r^2=0.473$, $p<0.0001$ & $r^2=0.331$, $p<0.0001$). Together, these results indicate that *Type IV* clones are enriched in primary and secondary metastatic samples. Overall, secondary metastatic sites are more similar to STG282CT-x3m than their paired non-metastatic replicate (STG282CT-x1_1 & STG282CT-x1_2).

We sought to better understand the number of specific clones shared between matched metastasis and primary STG282CT samples, *Figure 5.6.4* shows the degree of overlap in detected barcoded clones, stratified based on clonal phenotype (*Type I, IV & V*). In *Figure 5.6.4 A* we can clearly see the majority of overlap between samples in *Type I* clones is between non-metastatic sights. Each shared 59 barcodes with the other, and a total of only 20 with the metastatic samples. *Type IV* clones (*Figure 5.6.4 B*) show a great deal of overlap between all samples (134 clones) and few specific to metastatic (42 specifically shared, 49 unique to metastatic in total) or primary (14 specifically shared, 69 unique to primary in total) samples. *Type V* are slightly more evenly distributed, though still show 49 clones shared between primary sights and 23 with metastatic samples. Hence, metastatic deposits show consistent enrichment in the cellular content of *Type IV* clones. These clones may remain dormant in primary samples before seeding and expanding in metastatic sites.

Finally, we sought to rationalise the enrichment of clonal populations of a particular phenotype to metastatic samples. Aside from estimating the absolute size of a clone, in terms of cell number, the relationship between RBA and spiked in cell number derived in *Figure 5.3.3* allowed us to estimate the “clone initiation capacity” (CIC), *i.e.* the proportion of input cells of a given clone, able to establish a detectable clone of the same background upon passage. CIC as a fraction of input cell number was calculated and averaged across passages and expansion arms, results are shown in *Figure 5.6.5*. Each clonal phenotype was highly variable in CIC, this was as expected, as a great deal of clonal loss and/or fluctuation was seen during the course of these experiments. Perhaps also unsurprisingly, *Type IV* and *V* clones had significantly higher CIC than *Type I, II* or *III*. Both *Type II* and *III* clones are defined based on their downward trend and eventual loss in later passages, and they have a lower calculated ability to establish clones on passage. Indeed, the ability to do so defines both *Type IV* and *V* clones.

Figure 5.6.1 – Identification of metastasis in STG282CT

A) Shows a representative image of suspected micro-metastatic sites identified around the base of murine right lower lung lobule (black arrows), as imaged by the Xenogen small animal imaging system. These regions were dissected and stained with H&E for visual inspection. In each case where potential micro-metastatic sites were identified; H&E showed no obvious tumour structures. In total, we isolated and inspected >100 mammary glands and tumour draining lymph nodes without finding definite metastatic sites. **B)** We identified a <1 cm³ cell mass (arrow 2) at the base of the 4th (abdominal; arrow 3) mammary gland near the site of tumour fragment implantation (arrow 1). Though this mass was not physically attached to the tumour bulk, it could have resulted from cells becoming dislodged during implantation and may not in fact be a local metastasis. Nevertheless, the fragment was isolated and dissociated into six equally sized fragments, one for NGS and barcode analysis, five for reimplantation into NSG mice. **Figure Overleaf.**

Figure 5.6.1 – Identification of metastasis in STG282CT

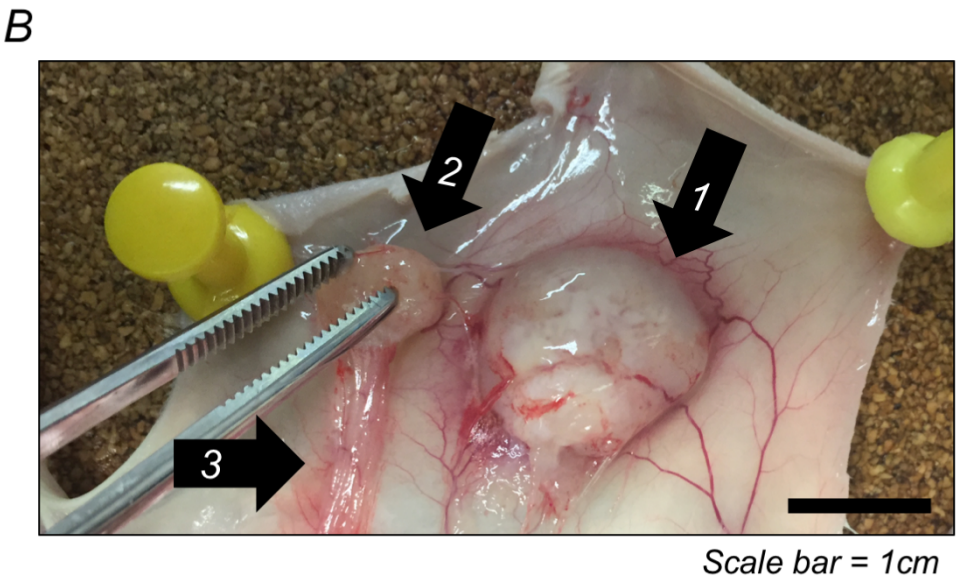
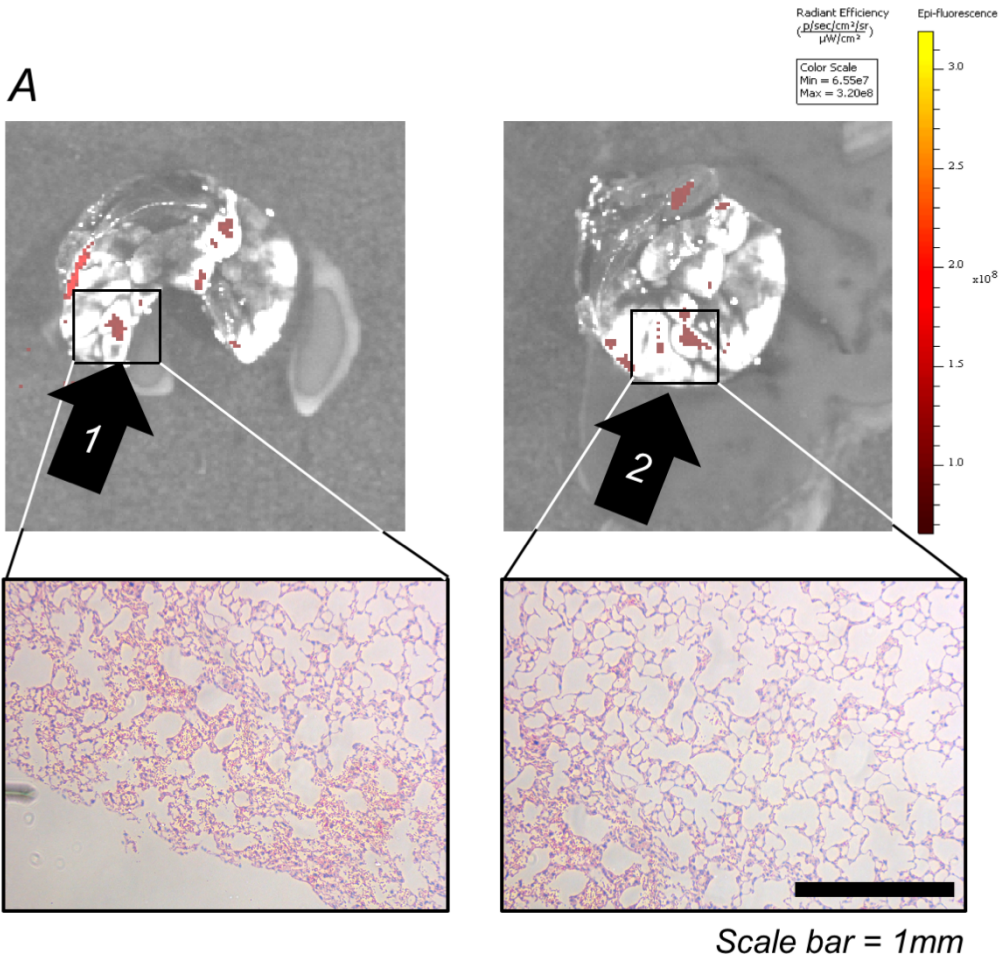


Figure 5.6.2 – Clonal phenotypes in STG282CT metastatic deposits

*Shows a schematic for expansion of this micro-metastasis (designated STG282CT-x3m) together with the results of ClonTracer barcode analysis. Barcodes seen in previous samples (~100% of those isolated) were assigned a clonal Type (I-V) based on repopulation dynamics of the barcode in Figure 5.5.1. Two out of five STG282CT-x3m progeny formed similar micro-metastasis in local mammary glands. **Figure Overleaf.***

Figure 5.6.2 – Clonal phenotypes in STG282CT metastatic deposits

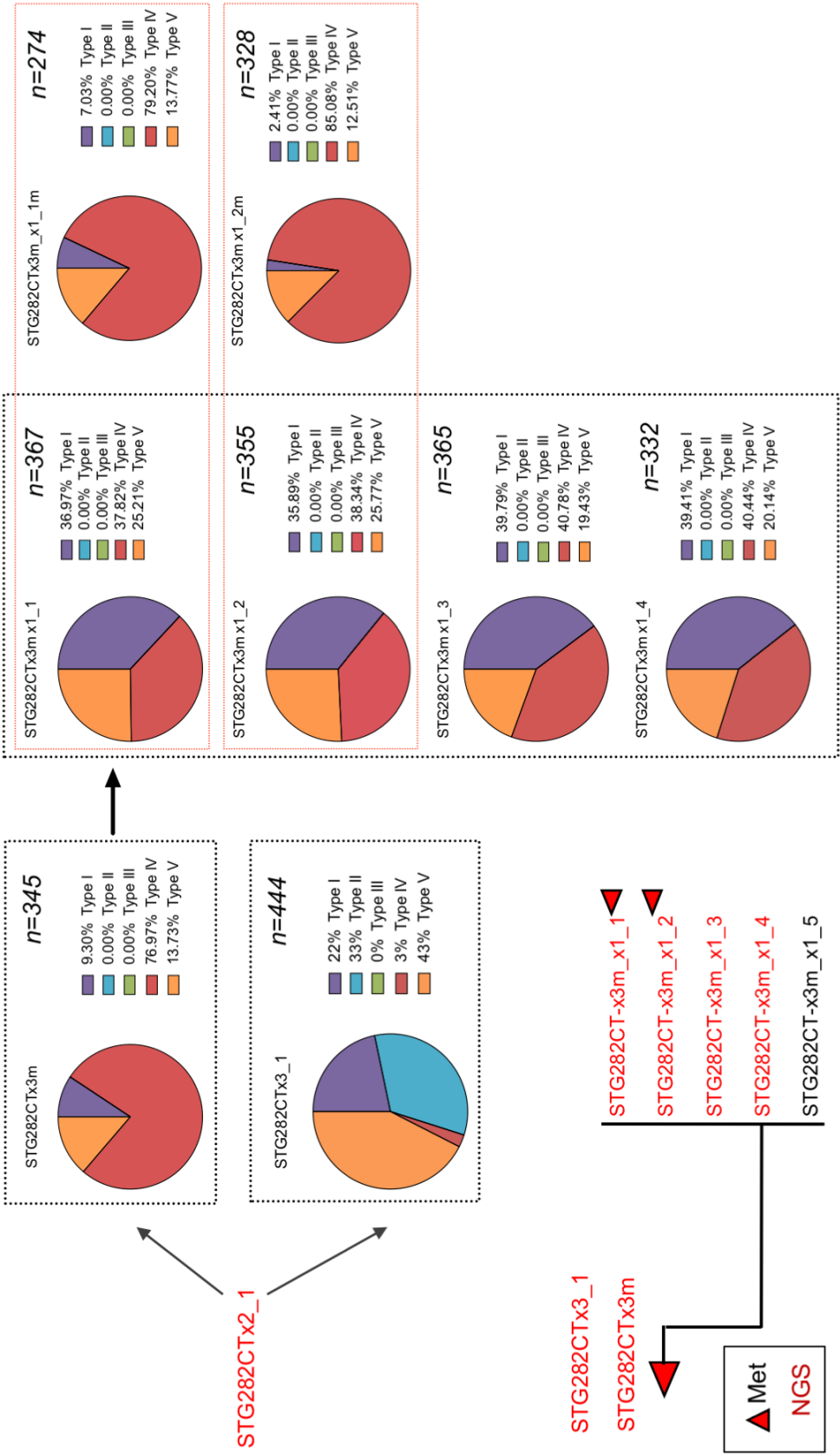


Figure 5.6.3 – Overlapping barcode abundance in metastases

Comparison of the relative abundance of 152 shared barcodes in paired metastatic and primary STG282CT models. **A)** Compares secondary metastases to their paired non-metastasis primary tumour. **B)** Directly compares RBA in secondary metastases. **C)** Shows relative barcode abundance in STG282CTx3m versus secondary metastasis sites (STG282CTx3m_x1_1m in the left-hand panel; STG282CTx3m_x1_2m in the left-hand panel). Points are coloured by clonal phenotype, in the same colour scheme as before (purple is Type I, red is Type IV, orange is Type V). **Figure Overleaf.**

Figure 5.6.3 – Overlapping barcode abundance in metastases

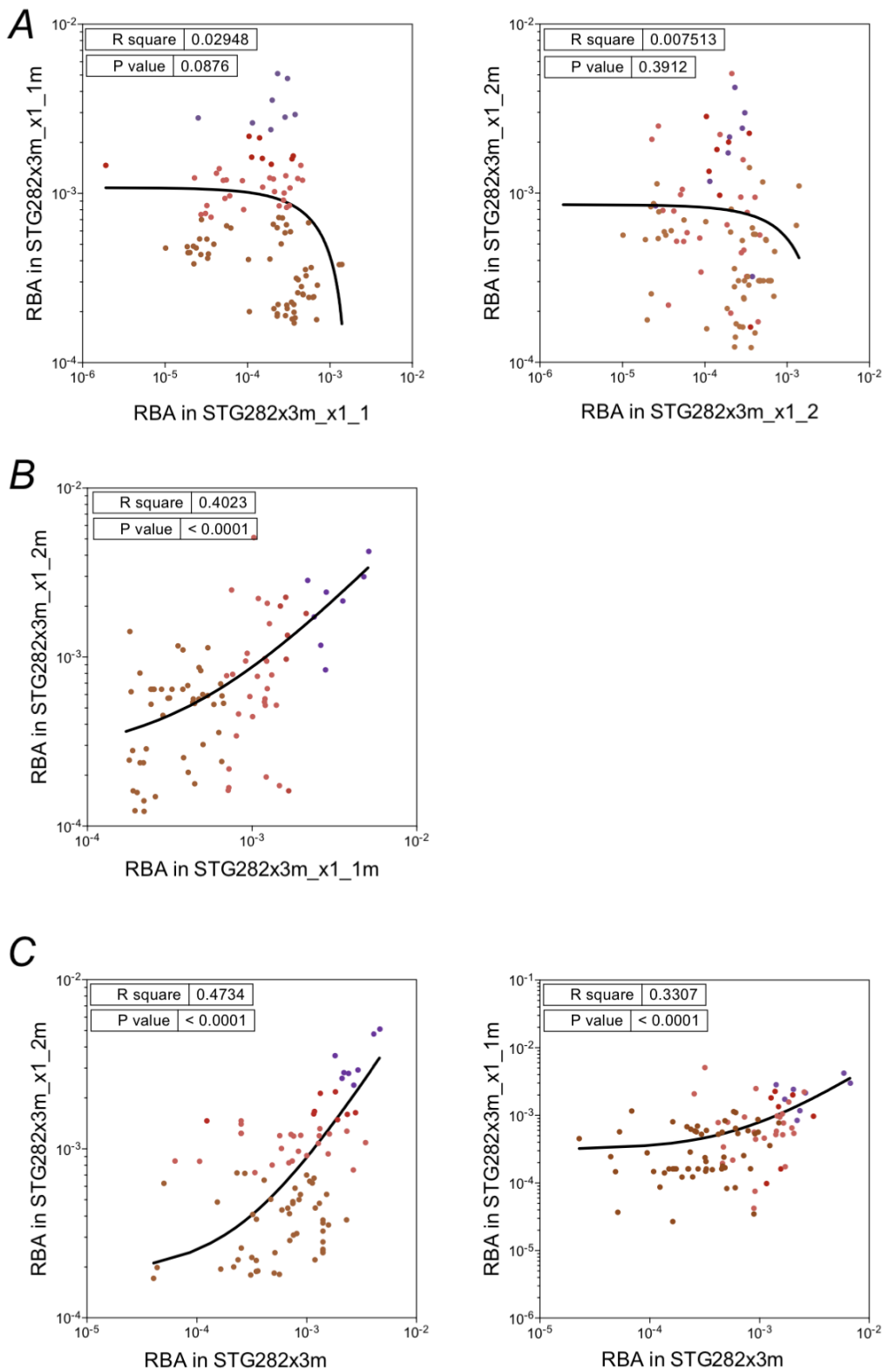


Figure 5.6.4 – Clonal phenotype representation in metastases

*Shows the degree of overlap in detected barcoded clones, stratified based on clonal phenotype (Type I, IV & V). **A)** shows Type I clones, **B)** Type IV clones and **C)** Type V clones.*

Figure Overleaf.

Figure 5.6.4 – Clonal phenotype representation in metastases

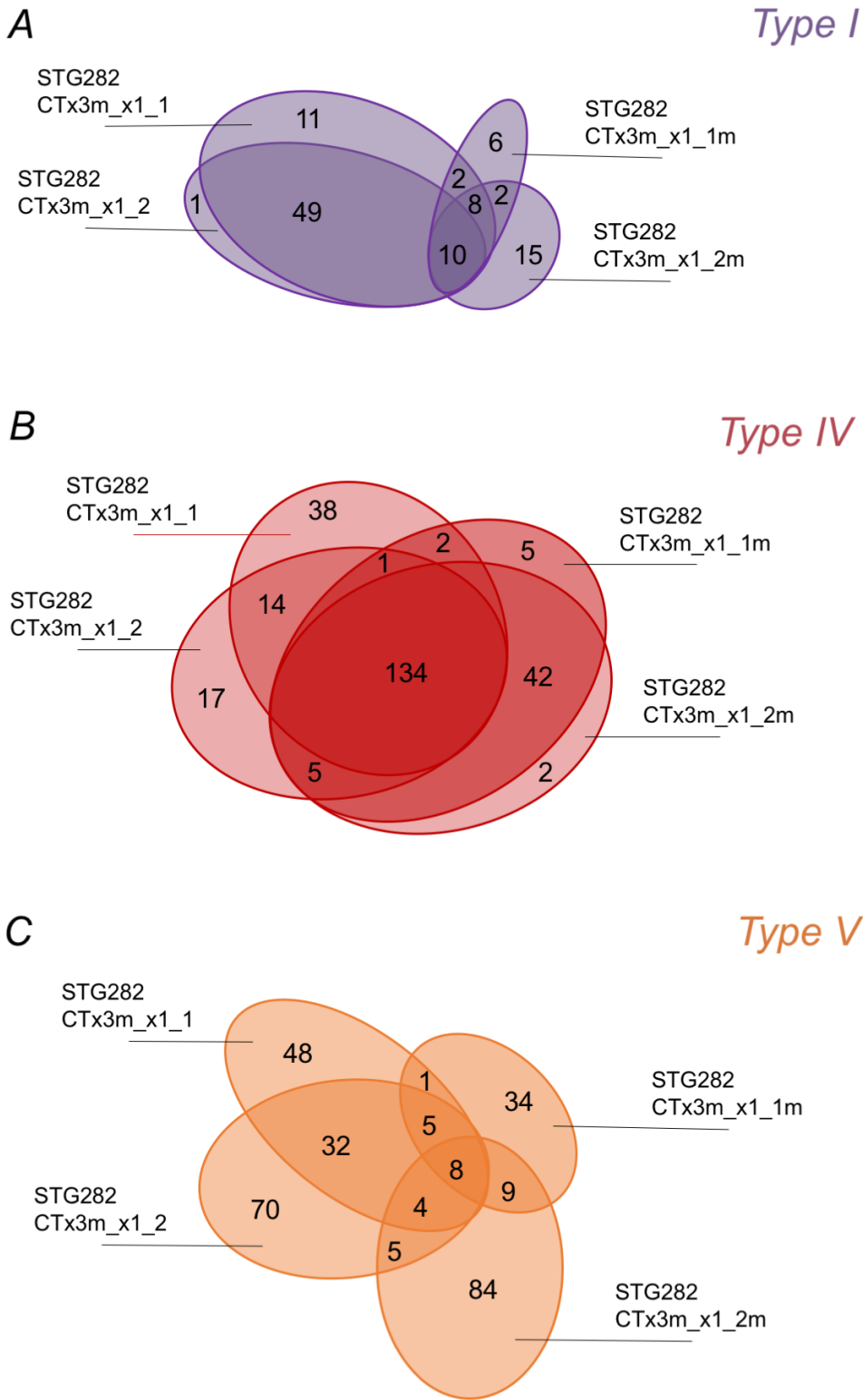
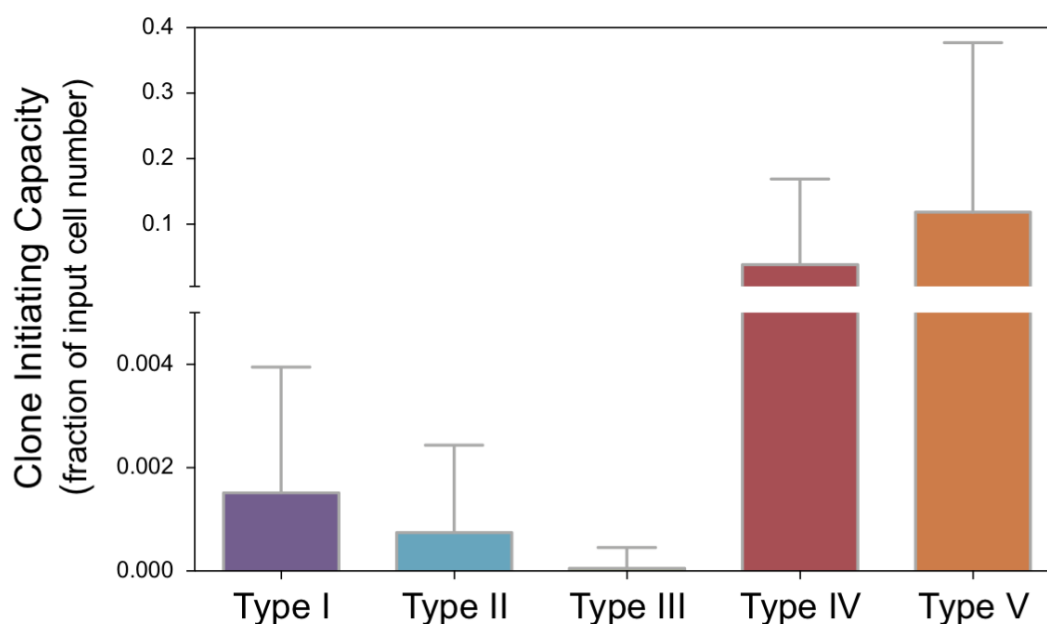


Figure 5.6.5 – Clone Initiating Capacity (CIC) of barcoded cells

Tukey's multiple comparisons test	Mean Diff.	95% CI of diff.	Significant?	Summary
Type I vs. Type II	0.0007708	-0.03039 to 0.03194	No	ns
Type I vs. Type III	0.001466	-0.03108 to 0.03401	No	ns
Type I vs. Type IV	-0.03684	-0.06972 to -0.003960	Yes	*
Type I vs. Type V	-0.1168	-0.1489 to -0.08479	Yes	****
Type II vs. Type III	0.0006951	-0.02922 to 0.03061	No	ns
Type II vs. Type IV	-0.03761	-0.06789 to -0.007332	Yes	**
Type II vs. Type V	-0.1176	-0.1470 to -0.08823	Yes	****
Type III vs. Type IV	-0.03830	-0.07000 to -0.006605	Yes	**
Type III vs. Type V	-0.1183	-0.1492 to -0.08746	Yes	****
Type IV vs. Type V	-0.08001	-0.1112 to -0.04881	Yes	****

Figure 5.6.5 – Clone Initiating Capacity (CIC) of barcoded cells

We used the relationship between RBA and spiked in cell number derived in Figure 5.3.3 allowed us to estimate CIC, i.e. the proportion of input cells of a given clone, able to establish a detectable clone of the same background in upon passage. CIC as a fraction of input cell number was calculated and averaged across passages and expansion arms, with statistical significance based on a one-way ANOVA with Tukey's multiple comparison correction given in the table below the figure. (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$).

5.7 General Discussion

We have recently described a biobank of breast cancer PDX models, with extensive molecular, genomic and clinical annotation (198). Through characterisation of the specific repopulation dynamics of genomic clones in these models, we add to a growing body of evidence showing PDX models largely reflect the complex molecular and architectural features of their originating patient tumour samples (23,307,445,446). Moreover, we have developed an integrated platform allowing the use of our PDXs in high throughput combinatorial drug screening, and offered extensive commentary on the use of these models to *i*) maintain and study intratumour heterogeneity (23) and *ii*) aid in the preclinical drug discovery process (289). The experiments described in this chapter provide a framework for the interrogation of cellular clonal populations in our PDX models through high resolution lentiviral lineage tracing. The ability to deconvolute cellular and genomic clonal dynamics in these highly complex models of human malignancy offers the potential to redefine our understanding of tumour heterogeneity (289). Reproducible patterns of clonal dynamics across paired transplantations, shown in this Chapter, suggest clonal kinetics may represent non-stochastic selection events. In *Chapter 3*, lentiviral lineage tracing technology was used to trace the origins of tamoxifen resistance in the MCF7 cell line. It is hoped that following these optimisation steps, lentiviral lineage tracing could be applied in treatment resistance studies across our biobank of PDX models.

In this Chapter, we established a methodology for lentiviral transduction of PDX models. In particular, we modified existing protocols to optimise for hyper concentration of viral supernatant using extensive ultracentrifugation with a sucrose gradient. Viral titres showed our supernatant was indeed highly concentrated in active virions (see: *Methods, Section 2.3*). Optimisation was primarily carried out with a biologically inert HIV-ZsGreen1 virus to avoid repeat freeze-thaw cycles or complete use of our validated ClonTracer lentiviral barcode library. PDX models were relatively resistant to

puromycin, and so we opted for an RFP based titration for ClonTracer virions and GFP based for HIV-ZsGreen1 (*Figure 5.2.1*). Puromycin is an amino nucleoside antibiotic produced by the bacterium *Streptomyces alboniger*. It inhibits protein synthesis by disrupting peptide transfer on ribosomes causing premature chain termination during translation. It is a potent translational inhibitor in both prokaryotic and eukaryotic cells. Resistance to puromycin is conferred by the puromycin N-acetyl-transferase gene (*pac*) from *Streptomyces*. Puromycin has a fast mode of action, causing rapid cell death at low antibiotic concentrations. Adherent mammalian cells are typically sensitive to concentrations of 2 to 5 µg/ml, while cells in suspension are sensitive to concentrations as low as 0.5 to 2 µg/ml (361,457–459). Hence, resistance to puromycin at 5 µg/ml in STG282 was higher than expected from reports in literature and we sought to avoid any potentially off target effects of using puromycin at high doses. Both HIV-ZsGreen1 (*Figure 5.2.2*) and ClonTracer (*Figure 5.2.3*) transduced STG282 cells were relatively slow growing when injected into the flanks of NSG mice. Certain primary human cells, particularly those that are quiescent in nature, are reproducibly challenging to transduce with lentiviral constructs (449,450). Clearly, quiescent (*e.g. Type IV clones* (196)) could hold particular importance in our later functional studies (451), and a molecular barcoding technique which selected for non-quiescent cells should be avoided. However, each second passage transduced tumour (*i.e. STG282-x3_ZsGreen1x1 & STG282CTx1*) grew at a rate not dissimilar to wildtype tumours. We reasoned delayed growth on implantation could, thus, be at least partially attributed to the process of flow sorting rather than selective infection of a slow growing sub-population of cells.

Delayed onset of growth, together with difficulties in achieving a high infection efficiency, led us to question whether we were selecting for a subpopulation of cells within our PDX models. Clearly, as we hoped to study the causes and consequences of tumour heterogeneity, any method which selected for a rare subpopulation should be avoided. At least partial preservation of heterogeneity was confirmed by profiling infected and parental PDX models by sWGS (*Figure 5.2.5*) for copy number and IHC for

protein expression (*Figure 5.2.4*). Further characterisation could have been performed by SNV profiling and/or PyClone reconstruction of clonal population. This work is ongoing.

We initially planned on increasing overall barcode representation through a system of implantation in multiple NSGs, tumour harvesting and pooling, followed by re-transplantation. In this way, we hoped to define two experimental arms, each containing PDX models with the same number and type of barcodes. A drug resistance screen could then be designed in a similar fashion to that described with MCF7s, described in Chapter 3. However, it quickly became apparent that significant clonal selection was occurring on PDTC implantation and this was precluding the development of a pool of barcoded cells with equal distributions.

Clonal selection and a progressive loss of clonal diversity in breast cancer PDX models has also been described in a study by *Nguyen et al.*, (197) using a similar approach as described in this Chapter, the authors found that the total number of detectable clones decreased from as many as 185 to just 12 in three passages. It was reasoned that this was due to different clone initiation frequencies of transplanted cells, which was found to vary from $\sim 1/10$ to $\sim 1/10,000$ (197). In our models, we found a decrease in overall clonal diversity, measure by the SDI, from an average of ~ 300 to ~ 100 -150 (*Figure 5.3.4*). Overall clone numbers dropped from ~ 1000 in passage x1 to ~ 350 in passage x5 models. It is important to note that this reduction in clonal diversity does not necessarily translate to a reduction in diversity of clonal phenotypes. The probability of any cellular clone contributing to any transplant is reflected in its relative distribution in the previous donor tumour; abundant clones have a higher probability of persisting than less abundant clones. Through reconstruction of genomic clonal structures (199), we have previously shown that the majority of outright clonal selection occurs on initial implantation, and that only $\sim 20\%$ of clonal populations show significant changes in cellular prevalence across several years of *in vivo* culture (198). It would be interesting

to further characterise the heterogeneity of cellular populations in our models through techniques such as single-cell RNAseq. Indeed, malignant cells can occupy diverse physiological states resulting from stress (hypoxia, DNA damage, starvation), quiescence, or cell cycle stage. Thus, heterogeneity may even exist across a single barcoded clone; advances in single-cell RNAseq may enable us to identify and describe complex phenotypic populations even between cells of the same barcode and genomic background (460).

We reasoned that by analysing barcode presence at different passages of our STG282CT model, we could study clonal selection and competition over time. Furthermore, by analysing their relative abundance we could derive a model of clonal fitness. *Figure 5.3.1* shows a schematic of the expansion of the STG282CT model. Model numbers in red represent those where NGS and barcode analysis was performed, and red triangles indicate models in which micro or local metastases were found. In *Figure 5.4.1* we follow the fate of individual barcodes over several passages, noting patterns of symmetric and asymmetric growth. In *Figure 5.5.1* we group barcode clones into clonal phenotypes defined by *Kreso et al.*, (196) and in *Figure 5.5.2* we examine the proportion of each clonal phenotype in each passage of our barcoded PDX models.

Conceptually, variation in clonal behaviours could be stochastic or they could reflect specific classes of breast cancer PDX cells that function in a predictable deterministic fashion. If the process is stochastic then the probability of any barcoded clone contributing to any transplant would be reflected in its relative distribution in the previous donor tumour; abundant clones would have a higher probability of persisting than less abundant clones. By contrast, the emergence of dominant barcoded clone from minor or undetected barcodes would be a low probability event, especially if this occurred simultaneously within multiple paired implantation events. Examination of clonal output as measured in parallel recipients at each passage, either by individual RBA or by proportion of clonal phenotype (*Type I, II, III, IV, V*), did not support the stochastic

model in our samples. Indeed, in the majority of paired implantations, we see symmetric growth patterns (*Figure 5.4.1*) coupled with similar proportions of clonal phenotypes (*Figure 5.5.2*). The expansion arm following STG282CT-x2_1 showed the most asymmetric repopulation dynamics. For example, in asymmetric growth in STG282CT-x3_b1 matched transplants (into STG282CT-x4_b1 and STG282CT-x4_b2; *Figure 5.4.1 B* & *Figure 5.5.2 B*) was shown to be due to the appearance of a group of clones not previously seen in any model (*Type IVs*). Moreover, asymmetric growth in further reimplantation of STG282CT-x4_b2 was attributed to the loss of *Type V* clones in only one model, rather than a repetition of the *Type IV* clone selection seen in the previous generation.

We noted that the framework by which clonal phenotypes are defined is entirely dependent on experimental design. For example, a *Type IV* clone could be classified as a *Type V* had the experiment consisted of a further passage (presuming it was undetected), likewise, a *Type I* or *II* clone could in fact be a *Type IV*. *Nguyen et al.*, defined several clusters based on the kinetics of repopulation in their breast cancer cell-line xenografts and PDX modes (197). In this case the authors relied on *k*-means clustering, a technique where *n* observations are partitioned into *k* clusters so as to minimize within-cluster variance (453). For each experiment, the authors thus defined several distinct kinetic behaviours, and in many cases these went beyond the five clusters proposed by *Kreso et al.*, (196). For example, in MDA-MB-231 derived xenografts three clusters were defined: i) a relatively stable clone size, ii) a relative clone size increasing on passage and iii) a relative clone size decreasing in the second passage. In a second MDA-MB-231 experiment, a further two clusters were defined, both characterized by delayed clonal growth. One cluster was further characterised by being detectable in secondary tumours but decreasing in tertiary tumours. SUM-149 derived xenografts, despite being expanded over a single passage *in vivo*, yielded six and seven clusters, though these were variations of the same neutral, positive and negative growth trends. PDX models analysed by the group shared only three of the patterns exhibited by the serially passaged cell lines; no

change, decreasing or fluctuating; but not clones that increased in size or showed delayed growth onset (197). Ongoing work will aim to classify clonal repopulation dynamics in an unsupervised fashion. This will likely take the form of a bespoke model-based classifier: k-means clustering is heuristic, meaning it is best suited for clustering Gaussian distributions, and may not be as accurate when grouping cellular growth patterns, which are typically exponential in nature (454).

Despite the noted limitations of our five clonal behaviours clusters, they did allow us to generate novel insights in three cases where suspected local metastases were seen. Specifically, *Figure 5.6.2* shows enrichment of specific *Type IV* clones in STG282CT-x3m. When this metastasis was re-implanted into four second generation models, we see an overall reduction in the proportion of *Type IV* clones, in line with the spread of clonal phenotypes typically observed in models of this passage. Of not, two of these second generation STG282CT-x3m models (STG282CT-x3m_1 & STG282CT-x3m_2) also showed local metastasis, and these metastases were again enriched for *Type IV* clones. Taking this analysis further, we find that both the secondary metastasis and their corresponding primary tumours share the majority of *Type IV* clones (*Figure 5.6.4*). Thus, the same *Type IV* clones are present in both primary and metastasis sites but lay dormant in the primary and dominate the metastasis. Moreover, across all passages we find *Type IV* clones, along with *Type V*, have significantly enhanced CIC relative to *Type I, II & III*.

Metastatic disease remains largely incurable and the main cause of cancer-related death across organ sites. Metastasis is thought to be the end result of a multistage process that includes local invasion by the primary tumour cells, intravasation into the blood or lymphatic system, survival in circulation (hematogenous and/or lymphatic), arrest at a distant organ, extravasation, survival in a new environment, and metastatic colonization. Each of these steps relies on specific phenotypic features of the tumour cell, as well as interactions with the host microenvironment and the immune system (151,152). There are two general models of metastatic dissemination: the linear progression model and

the parallel progression model (150). Both models assume that the primary tumour and its metastases are clonally related, in that they derive from a common ancestral cell. In the linear progression model, metastases emerge from late occurring advanced clonal subpopulations (153). The parallel model suggests that a metastasis is seeded early in molecular time from the primary site and progresses through the independent acquisition of mutations which may be different to the primary (124). A third mechanism, the concept of metastasis-to-metastasis seeding, the cascade hypothesis, was proposed in 1975, but until recently has remained somewhat under studied (154,155). This describes a process by which polyclonal seeding may result from a combination of clones from both the primary and other metastatic sites, as has been described in cases of lethal metastatic prostate cancer (461,462).

In our models, we can consider similarity to the originating sample as indicative of parallel progression (as both metastasis and primary are seeded by the same proportion of cells) and similarity to the matched primary as indicative of linear progression (as metastasis is seeded from a late stage primary with the corresponding late stage primary composition of barcodes). In *Figure 5.6.3* we see each metastatic site has a higher degree of concordance with the originating sample than the matched primary. This could be indicative of parallel seeding of metastasis; however, we see an enrichment of *Type IV* clones in all three cases of metastasis, but not primary PDX tumour derived from metastasis, suggesting at least some functional clonal selection is occurring. These conclusions, together with other indications of clonal selection and mirrored evolution in our models, would be greatly aided by genomic analysis. Further implementations of the barcoded PDX program are likely to make use of paired genomic and cellular clonal analysis, as we have carried out and presented in *Chapter 3*.

In summary, this chapter presented optimisation and establishment of a TNBC PDX model with a methodology for clonal tracking at a resolution of >500 cells. Although clonal diversity was reduced on serial transplantation, we have also shown minimal

asymmetric clonal kinetics in paired transplantation assays. We also define clonal phenotypes (*Type I-V*) and find quiescent, *Type IV*, clones are enriched in metastatic deposits. Further work will include genomic analysis in these models and expand the breadth of barcoded models to encompass the high degree of heterogeneity in human breast cancers represented by our PDX biobank (7,198).

Chapter 6

Summary, Conclusions & Future Work

- Context for this work: Breast Cancer Incidence & the Clinical Importance of Tumour Evolution
- Resistance to Endocrine Therapy
- Clonal Tracing in Complex Preclinical Models

6.1 Context for this work: Breast Cancer Incidence & the Clinical Importance of Tumour Evolution

The estimated global incidence of all cancer types in 2015 was 17.5 million (463). Fourteen percent of all deaths in 2005 were due to cancer, which increased to 16% in 2015 (464). Breast cancer is the most common cancer overall, with an estimated 2.4 million new diagnosis in 2015; 1 in 14 women and 1 in 603 men will develop breast cancer between birth and age 79 years. The vast majority of breast cancers occur in women, with 2.4 million cases versus 44,000 cases in men. Breast cancer was the leading cause of cancer in all sociodemographic index (SDI) quintiles except for the high and high-middle SDI quintiles where it was the second most common cancer after prostate and lung respectively (463). In England, breast cancer accounted for 44,000 new cancer registrations in 2016 and was the most common diagnosed cancer (465).

As a quantitative measure of the burden of disease, a disability adjusted life year (DALY) can be thought of as one lost year of healthy life and is computed as the sum of years of life lost (YLL) due to premature mortality and the years lost due to disability (YLD) (465). Between 2005 and 2015, many countries experienced a decrease in cancer mortality despite increasing incidence rates. However, as of 2015, breast cancer accounted for 15.4 million DALYs in 2015, of which 88% came from YLLs, and 12% from YLDs (465).

These observations reinforce the global burden due to breast cancer. A better understanding of breast cancer biology coupled with advances in prevention, detection and treatment will become increasingly important over the next several decades. This includes characterising the origins and development of breast cancer in individuals with the expectation that a greater fundamental knowledge of tumour biology will lead to improved outcomes through the development of new clinical strategies.

Tumour evolution has been a key conceptual framework in cancer biology since it was first put forth by Peter Nowell in 1976 (12). The theory postulates that cancers arise from a single cell that has a selective advantage over its neighbours and that cancer can be understood based on the evolutionary principles of selection and adaptation originating from this ancestral cell. Over time, cells within the tumour continue to adapt and bestow on the tumour whole, specific traits described as the Hallmarks of Cancer (30,31). These ideas have been developed using many of the concepts first established in evolutionary biology (383,466); considering cancer as a disease of multicellular organisms in constant balance between Darwinian selection acting on the level of a single cell and the need for coordination between multiple cells for the good of the organism (467,468). From this perspective, cancers occur when an individual cell behaves in an autonomous manner, escaping from the mechanisms in place to coordinate cell behaviour (469).

The classic model of carcinogenesis describes multiple, successive clonal expansions driven by the accumulation of genomic changes or ‘mutations’ that are preferentially selected by the tumour environment (470). However, it is important to note that natural selection acts on phenotypes rather than genotypes. Indeed, selection can be transient, favouring a specific phenotype in response to fluctuating changes in microenvironment. Indeed, recent work has uncovered monogenomic clonal expansion of phenotypic clones responsible for tamoxifen resistance in breast cancer (376) and chemotherapeutic resistance in CRC PDX models (83,192). More broadly, tumour evolution and resultant heterogeneity have been linked to several clinically important facets of breast cancer (11,456,471), and further study to understand the causes and consequences of tumour evolution is warranted.

6.2 Resistance to Endocrine Therapy

Breast cancer is one of the few tumour types in which molecular classification (based on the presence of hormone receptors (ER α /PR) and HER2) has successfully been used for the design of individualized therapies, leading to significant improvements in disease-specific survival (472). However, despite the success of anti-endocrine therapies in ER α + breast cancer the development of treatment resistance is commonplace, and ultimately around 40% of patients will relapse during or after the completion of adjuvant endocrine therapy (375). Resistance to the endocrine therapy, and WHO essential medicine (318), tamoxifen, is thought to be due to differential ER chromatin recruitment by FOXA1. Resistance has variable been attributed to rapid reprogramming of FOXA1 binding on a short time scale (238), and the expansion of phenotypic clonal populations marked by active FOXA1 enhancer (376).

In *Chapter 3*, we utilise the ClonTracer lentiviral lineage tracing system (334) to uniquely tag over 1 million MCF7 cells and follow their clonal behaviours during the development of tamoxifen (4-OHT) resistance. We find that resistance is associated with the selection of a subset of cellular clones, genomically similar to those sensitive to therapy. Although we find remarkably similar clones are selected across replicates, the resultant tamoxifen resistant sublines are characterised by several differentiating features: differences in proliferation rates, ER α -signalling and resistance to rapamycin are noted. Future work in this regarded should focus on an unbiased screen of the mechanisms by which resistant cell lines differ. An RNA-seq based classification of resistant lines is currently underway and should offer insight into regulatory pathways involved in the specific mechanism of tamoxifen resistance in our models. Furthermore, the addition of further time points in our NGS analysis would afford greater resolution in the description of clonal phenotypes in our control clones. For example, barcoded clones could be classified based on their passage-kinetics, in a similar manner to *Kreso et al.*, (83), this would

enable further scrutiny of our hypothesis that quiescent, or slowly dividing, MCF7s are responsible for the development of tamoxifen resistance.

Recently, *Hinohara et al.*, barcoded MCF7s using the ClonTracer system (135) during the development of both tamoxifen and fulvestrant resistance. The authors found that resistance to each drug was due to the selection of a rare, pre-existing, population of cells (483). Utilising a mathematical model of stochastic population dynamics, the authors were able to estimate the fraction of pre-existing barcodes in each scenario. Using an expected mutational frequency of 10^{-5} per division, the authors demonstrated resistant clones at 0.5-1% (for fulvestrant) and 1% (for tamoxifen) frequency in the initial population. These fractions are consistent with our own calculations. However, contrary to our work, the authors found genomic heterogeneity in resistant populations by exome sequencing. Gene set enrichment analysis showed that expression of genes downstream of some of the identified genetic variants were significantly altered and that associated expression signatures were also present in metastatic sites of those relapsing on tamoxifen (483). Given the genomic instability of MCF7s, coupled with the near extinction events in our drug treatment arm, it would be interesting to model whether any slight changes in VAF between naïve and resistant MCF7s would indeed be statistically significant.

Importantly, DNA hypomethylation at oestrogen-responsive enhancers has been associated with the development of ER α + breast cancers, and the reversal of this process with the development of resistance to endocrine therapy (285). Moreover, heterogeneity amongst tamoxifen resistant models derived in parallel processes has been studied previously (345) and the resistance process has been linked in the literature to pluripotency factors such as SOX2 (277) and OCT4 (438,473). Taken together, tamoxifen resistance is broadly due to differential enhancer methylation and ER α chromatin interaction governed by pioneer factor recruitment. Parallels between this process and that of differentiation during embryonic development, coupled by reports of

the importance of embryonic transcription factors in the process, led us to question the role of CSC and/or dedifferentiation in our models.

In *Chapter 4*, we study the role of OCT4 in M-ICs and the development of tamoxifen resistance. Though resistance to 4-OHT was the key phenotype under investigation, there is evidence that our resistant lines were oestrogen-independent and may have shown resistance to AIs *in vivo*. We find that the M-IC and tamoxifen resistant phenotypes share a great deal of overlap and that both are enriched for *POU5F1* (OCT4) and *SOX2* expression. Indeed, we find a great deal of overlap between barcoded clones enriched in our resistant studies and those in mammosphere cultures, suggesting the same population of cell may be responsible for both phenotypes. We also uncover a mechanism by which tamoxifen bound ER α may directly drive the expression of *POU5F1*. Although backed by genome wide ChIP-seq studies (408), this should be further confirmed by, for example, ChIP-PCR experiments in our models. Through knock-in and knock-down experiments, we further profile the role of OCT4 in driving a M-IC and BC-TIC phenotype in MCF7s. Finally, we find *POU5F1* is prognostic, in the METABRIC cohort, of poor disease-specific outcome in ER α + patients treated with endocrine therapies.

Intreigingly, a recent study by *Bhatt et al.*, may suggest an alternative mechanism for OCT4's interation with oestrogen signalling (282). Specifically, the authors performed *in silico* motif-enrichment analyses within the ER-binding peaks in response to E2 or 4-OHT, to identify factors that would specifically recruit ER to genomic binding sites in the presence of 4-OHT as compared to E2. The authors found, and verified by ChIP analyses, Nkx3-1 and Oct-transcription factor homodimer motifs were enriched in 4-OHT-ER α preferential binding sites. Functionally, and like our own studies, the authors find OCT4 expression increased in TamR cell lines and essential to both tamoxifen resistance and MCF7 tumour growth in the presence of 4OHT *in vivo*. The authors conclude that OCT4 is a novel ER α coregulator, responsible for tamoxifen speicifc gene transcriptional

programs which ultimately lead to resistance (282), presumably through FOXA1 reprogramming or similar mechanisms.

A great deal of further work could be undertaken to uncover the true functional role of OCT4 in breast cancer and epigenetic reprogramming leading to tamoxifen resistance. In addition to uncovering related signalling pathways (related to the p38MAPK-Skp2-Nkx3-1 axis or otherwise), 4-OHT-ER α or OCT4 ChIP experiments would shed further light on regulatory mechanisms. Further, applying the ClonTracer labelling methodology to *POU5F1* overexpressing cells as they develop tamoxifen resistance, could help determine whether OCT4 is a true enabling factor in the resistance process. Furthermore, the mechanisms by which OCT4 contribute to transcriptional reprogramming could be uncovered through whole genome epigenomic mapping in knock-in/out cell lines.

6.3 Clonal Tracing in Complex Preclinical Models

In order to best study the causes and consequences of tumour evolution and resultant tumour heterogeneity, we must develop and characterise pre-clinical models that better reflect this heterogeneity (23). We have recently published a well annotated biobank of human breast cancer PDX models and adapted their use for high throughput drug screens (198). We find PDX models are relatively genomically stable over multiple *in vivo* passages and accurately reflect the molecular and histological features of their originating samples (195,198). In *Chapter 5*, we develop and characterise a TNBC PDX model tagged with the ClonTracer lentiviral lineage tracing system (334).

Through multiple rounds of *in vivo* passage, we find reproducible engraftment kinetics across thousands of cellular clones in multiple rounds of replicate engraftments. Marked populations ranged in size from 500 cells (our limit of accurate detection) to more than 1.6×10^5 cells. Like *Nguyen et al.*, we find that multiple rounds of passage generally leads to a reduction in the number of clonal populations detected and that clones vary in their CIC (197). By following a framework set out by *Kreso et al.*, we classify clonal populations into five distinct types on the basis of their repopulation kinetics (83). We find that relatively quiescent *Type IV* clones dominate in local metastasis seen in our models, and that on re-implantation a metastatic phenotype linked to these *Type IV* clones is selected. Although when previously profiled by NGS the STG282 model was relatively genomically stable on passage (195,198), future work in this regard should apply genomic clonal clustering techniques, such as PyClone (199), specifically to our barcoded STG282CT model. Genomic clonal selection in our identified metastatic sites will be of particular interest. Moreover, such analyses will help us understand any clonal (or cell-type) selection artefacts introduced by our lentiviral infection protocols. As a model system particularly suited to preserving intratumour heterogeneity, it is imperative that selection of subpopulation of cells by selective pressures associated with flow cytometry or lentiviral infection should be avoided.

Interestingly, a recently study from *Echeverria et al.*, (484) reported the use of the Collecta barcode system in a panel of TNBC PDX models to monitor the fate of individual tumour cell lineages during treatment with Adriamycin combined with Cytosine (484). The authors found that PDXs initially exhibited partial sensitivity, followed by maintenance of residual tumours that were resistant to chemotherapy and eventual relapse of a partially sensitive tumour. Interestingly, the authors report that residual tumours maintained the same clonal architecture as untreated tumours, but that a shared population of barcoded cells (representing ~20% of the total pool) were responsible for repopulation of the tumour following discontinuation of treatment. Further, exome sequencing revealed that mutant allele frequencies were largely conserved during the same process. The authors go on to suggest that residual tumours exist in a distinct state characterized by alterations in EMT, metabolic, and cell adhesion programs during treatment, but that these phenotypes reverted on continuation of tumour growth. *In silico* prediction of drug sensitivity, verified by pharmacological studies, revealed several potentially druggable drivers of resistance in the residual tumour state, offering the intriguing possibility of a specific treatment window to inhibit tumour regrowth following the cessation of chemotherapy (484).

Although optimised in a single TNBC PDX model, our own expansion of the ClonTracer system across a comprehensive PDX biobank is ongoing. Aside from the case of tamoxifen resistance, studied extensively in *Chapter 3* and *Chapter 4*, the ability to track the fate of individual cellular clones, perhaps in tandem with genomic clonal reconstruction, will be of great benefit in the study of multiple cases of treatment resistance in breast cancer. For example, RAD51 foci has recently been identified as a functional biomarker of homologous recombination repair and PARP inhibitor resistance in BRCA-mutated breast cancer (474). In one of our PDX models, PARP inhibitor resistance could be reverted on combination therapy with an ataxia-telangiectasia mutated (ATM) inhibitor. Such cases offer clinically relevant case studies

where high-resolution cellular tracking could significantly aid in our understanding of the development of drug resistance.

Appendix

Publications Related to this Thesis

- **Cassidy JW**, Caldas C, Bruna A. Maintaining Heterogeneity in Patient Derived Tumour Xenografts. *Cancer Res.* 2015;75(15):1–6. **Included.**
- Bruna A, Rueda OM, Greenwood W, Batra AS, Callari M, Batra RN, **Cassidy JW et al.** A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. *Cell.* 2016 Sep;167(1):260–274.e22.
- **Cassidy JW**, Batra AS, Greenwood W, Bruna A. Patient-derived tumour xenografts for breast cancer drug discovery. *Endocr Relat Cancer.* 2016 Dec;23(12):T259–70.
- **Cassidy JW**, Bruna A. Chapter 4 - Tumor Heterogeneity. Patient Derived Tumor Xenograft Models. In Academic Press; 2017. p. 37–55.

Review

Cancer
Research

Maintaining Tumor Heterogeneity in Patient-Derived Tumor Xenografts

John W. Cassidy, Carlos Caldas, and Alejandra Bruna

Abstract

Predclinical models often fail to capture the diverse heterogeneity of human malignancies and as such lack clinical predictive power. Patient-derived tumor xenografts (PDX) have emerged as a powerful technology: capable of retaining the molecular heterogeneity of their originating sample. However, heterogeneity within a tumor is governed by both cell-autonomous (e.g., genetic and epigenetic heterogeneity) and non-cell-autonomous (e.g., stromal heterogeneity) drivers. Although PDXs can largely recapitulate the polygenomic architecture of human tumors, they do not fully account for heterogeneity in the tumor microenvironment. Hence, these mod-

els have substantial utility in basic and translational research in cancer biology; however, study of stromal or immune drivers of malignant progression may be limited. Similarly, PDX models offer the ability to conduct patient-specific *in vivo* and *ex vivo* drug screens, but stromal contributions to treatment responses may be under-represented. This review discusses the sources and consequences of intratumor heterogeneity and how these are recapitulated in the PDX model. Limitations of the current generation of PDXs are discussed and strategies to improve several aspects of the model with respect to preserving heterogeneity are proposed. *Cancer Res*; 75(15): 2963–8. ©2015 AACR.

Introduction

Despite remarkable advances in our understanding of the progression of human malignancies and the molecular events that underpin tumor survival, new therapies often fail to show significant efficacy in clinical trials. Projects such as The Cancer Genome Atlas and METABRIC have demonstrated the remarkable heterogeneity across tumors previously believed to be of the same subtype (1). It could be argued that clinical trials fail to sufficiently stratify patients based on relevant biomarkers of drug response: the response rate of an unscreened population to a molecularly targeted therapy typically lies between 10% and 20% (2). Patient stratification based on molecular determinants of drug efficacy and tumor heterogeneity allows for significantly greater responses—exemplified by the success of ALK kinase inhibitors in EML4-ALK-positive non-small cell lung cancers (3). However, even with patient stratification, clinical responses can be fleeting, often adding only 6 to 12 months before disease progression (2). Thus, understanding intertumor heterogeneity is the first step toward improved drug efficacy and diminishing tumor relapse.

Intratumor heterogeneity is governed by both cell-autonomous (e.g., genomic and epigenomic heterogeneity) and non-cell-autonomous (e.g., stromal heterogeneity) factors. This heterogeneity has clinical implications in patient-specific responses to therapy and the rapid emergence of resistance to targeted therapies (4). By capturing intra and intertumor heterogeneity, patient-

derived tumor xenograft (PDX) models have a clear advantage over traditional models, supporting their use in oncologic drug discovery and preclinical development. PDX models largely recapitulate cell-autonomous drivers of heterogeneity: exhibiting genomic clonal dynamics reminiscent of their originating tumor sample (5, 6). Moreover, phenotypically distinct isogenic cellular clones have been shown to drive resistance to chemotherapy in colorectal PDX models (7). The tumor microenvironment has long been known to play an essential role in tumor progression and its role in drug response is becoming apparent (8, 9). Although PDXs retain the 3D architecture found in human tumors, stromal and immune interactions may be altered by inter-species compatibility and cellular component deficiencies in host models.

The poor performance of so many investigational drugs suggests that preclinical tumor models lack clinical predictive power. Indeed, one of the most often cited reasons for clinical failure is a lack of preclinical models that recapitulate the complexity of human cancers. It is with this in mind that many research and pharmaceutical groups have turned to PDX models (10, 11). The establishment and predictive power of PDXs have been reviewed recently elsewhere (11). This review will focus on the limitations of current PDX models and how these can be addressed in the future, specifically in terms of maintaining the heterogeneous nature of human cancers (summarized in Table 1). PDXs are arguably the best models of tumor heterogeneity, and therefore perhaps the most powerful tools for investigating tumor biology. However, they may fail to fully account for many non-cell-autonomous drivers of heterogeneity (Fig. 1A), and should be adapted if they are to reach their full potential as predictors of clinical efficacy in cancer drug development.

Clonal Dynamics and Tumor Heterogeneity in PDX Models

Through the course of tumor initiation and progression, cancerous cells undergo repeated mutational events that may

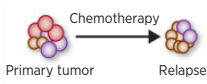
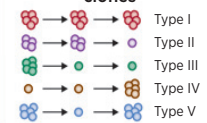
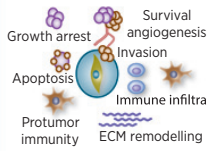
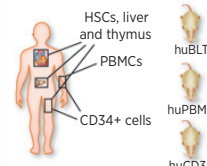
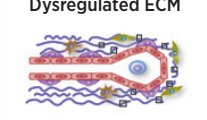
Breast Cancer Functional Genomics, Cancer Research UK Cambridge Institute, Department of Oncology, University of Cambridge, Li Ka-Shing Centre, Cambridge, United Kingdom.

Corresponding Author: Alejandra Bruna, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, United Kingdom. Phone: 44-1223-769667; Fax: 44-1223-769510; E-mail: alejandra.bruna@cruk.cam.ac.uk

doi: 10.1158/0008-5472.CAN-15-0727

©2015 American Association for Cancer Research.

Table 1. Sources of tumor heterogeneity, their consequences in translational and basic cancer biology, and how they are currently represented in PDX models

Source of heterogeneity		Implications for basic and translational research	Representation in current PDX models	Future prospects
Cell-autonomous	Genomic clonal dynamics 	Minor <i>KRAS</i> subclones predict resistance to EGFR targeted therapies in colorectal cancer. Overall clonal diversity correlates with drug resistance in ovarian and oesophageal cancers. ^{16,17}	Genomic clones reconstructed in a panel of 15 breast cancer PDX models revealed ongoing clonal dynamics. Polyclonal engraftment was possible, but clonal selection was clearly evident. ⁵	International collaborations such as the EuroPDX Consortium should facilitate sharing of expertise and eventually lead to increased engraftment with less pronounced clonal selection. ¹⁰
	Epigenetic/cellular clones 	Epigenetic 'attractor states' increase the phenotypic heterogeneity within the tumor and hence widen the pool of cellular clones able to contribute to treatment resistance. ²⁰	Colorectal and breast PDX models show five distinct cellular clonal phenotypes. ^{6,7} 'Type IV' quiescent clones were responsible for resistance to chemotherapy in colorectal cancer. ⁷	PDX models which better reflect the native tumor microenvironment should allow for more appropriate epigenetic clonal diversity.
Non cell-autonomous	Stromal heterogeneity 	Pro- and anti- tumor properties are attributed to different populations of cancer associated fibroblasts (CAF). Heterogeneous CAF or MSC populations could confer heterogeneity on the tumor bulk. ²⁶	Human stromal components are replaced by murine equivalents on PDX passage. It is unclear how closely mouse fibroblasts mimic their human counterparts in supporting tumor growth. Human fibroblast cell lines are heterogeneous in their ability to promote treatment resistance. ^{26,27}	Patient matched stromal components should be sourced whenever possible. Although human fibroblasts can be expanded <i>in vitro</i> , cell sorting may reduce engraftment efficiency of tumor cells and this should be considered.
	Immune infiltrate 	The role of the immune infiltrate on tumour progression is highly complex. Checkpoint inhibitors and other immunotherapeutics are promising new treatment strategies in oncology.	Highly immunodeficient mice are used for PDX implantation. The NSG strain is characterised by a lack of mature lymphocytes, the absence of functional NK cells, defective macrophages and defective dendritic cells. ³³	HuPDX immune models remain a substantial technical challenge. But the implications for study of tumor biology are profound.
	Dysregulated ECM 	Regulated ECM maintains tissue architecture and stem cell compartments. Loss of structure in cancer could contribute to oscillation between distinct transcriptional programs. ²¹	Matrigel is currently used to increase engraftment efficiency. Growth factors present in this murine basement membrane extract could support preferential engraftment of specific cell types. ECM is tissue specific, however in PDX models, ectopic implantation is commonly used. ^{23,25}	The ECM is tissue specific and orthotopic models should be considered where possible. Synthetic human alternatives to Matrigel should be investigated.

© 2015 American Association for Cancer Research

Cancer Research Reviews

AACR

NOTE: Strategies to improve the model, by better representation of both cell-autonomous (genomic and epigenomic clones etc.) and non-cell-autonomous (stroma, immune infiltrate etc.) drivers of heterogeneity are proposed.

or may not result in increased fitness relative to neighboring cells. Dramatic increases in fitness are seen with the acquisition of key driver mutations early in a tumor's evolution, for example the loss of *TP53* may lead to clonal dominance. However, selection operates on phenotypes in response to stress-inducing events, which may be stable or transient (12). A gain of fitness in one clone relative to another does not necessarily imply the loss of the latter in favor of the former. Rather, clonal populations within the tumor exist dynamically in space and time: competing, and perhaps cooperating, to further increase fitness of the tumor population as a whole (13). Clonal dynamics derived from tumors' inherent heterogeneity are thus extremely complex and can play key roles in tumor progression and development.

This conceptual framework of clonal evolution in cancer predicts several clinically observable features (14). First, every mutation or copy number aberration (CNA) present in the bulk tumor need not be present in all cells; indeed, spatial variation exists in a tumor's clonal composition. The existence of multiple subclones explains variable response rates to therapy, even within a single tumor mass, and the rapid emergence of drug resistance. For instance, the presence of a minor *KRAS*-mutant clone can predict colorectal cancer patients who will develop resistance to epidermal growth factor receptor (EGFR)-targeted therapy (4).

Our group has previously shown that breast cancer consists of at least 10 distinct molecular subtypes with significant differences in disease outcome and treatment responses (1). Furthermore, we have helped delineate the diverse, variable clonal composition of

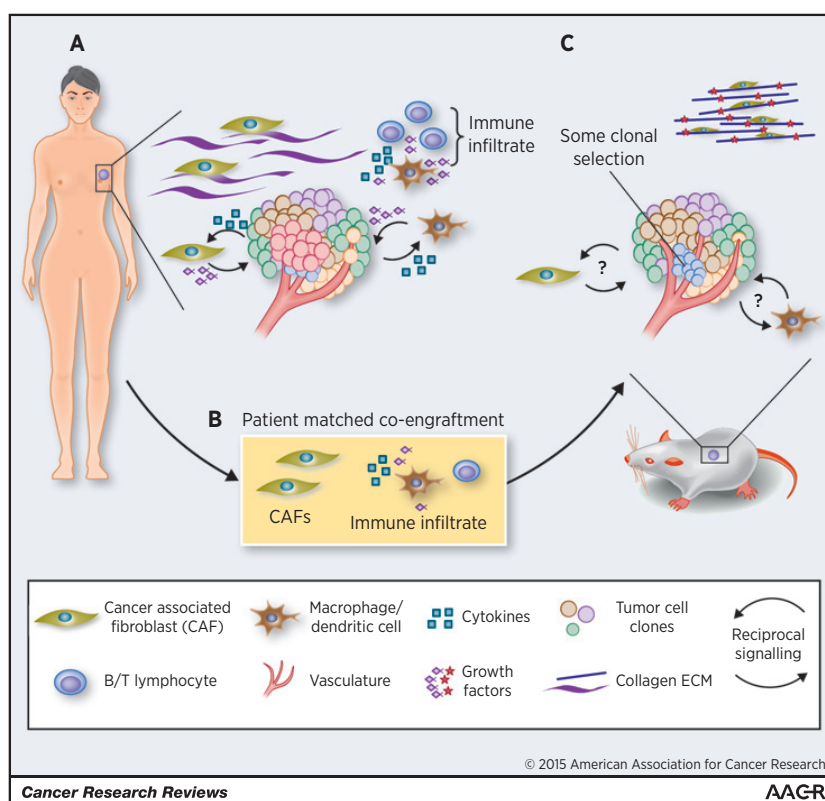


Figure 1.

A, a primary breast tumor including some sources of heterogeneity found in the native microenvironment. Reciprocal signaling pathways between tumor cells and TAMs and CAFs are highlighted. ECM is shown as collagen fibers with associated fibroblasts and macrophage/dendritic cells with T/B lymphocytes are shown as part of the immune infiltrate, though other cell types (NK cells, myeloid-derived suppressors, etc.) have been omitted for simplicity. B, patient-matched fibroblasts and immune cells as possible candidates for coengraftment in the next generation of PDX models. Current PDXs established in NSG mice lack an adaptive immune system and may have impaired innate immune cell infiltrates and cytokine signaling due to defective IL2 receptor (33). CAFs are known to contribute to treatment response; although murine fibroblasts are present in PDX models, it is unclear how faithfully these recapitulate their human counterparts (26). C, a PDX tumor in its native microenvironment. Questions over whether pro- and antitumor CAF/TAM signaling pathways are present to the same extent in PDX models as in the primary tumor are highlighted. Stromal and tissue architecture can have profound effects on transcriptional regulation but are often overlooked in the establishment of PDX models (25). To highlight potential differences in ECM organization between the native microenvironment and that of the PDX, here the ECM is shown as highly organized collagen fibers with closely associated myofibroblasts.

triple-negative breast cancers (TNBC; ref. 15). By allelic frequency measurements of 2,414 somatic mutations in 104 TNBCs, a complete spectrum of molecular and clonal compositions was characterized at diagnosis. Aside from the prognostic features of specific rare subclones (4), there is an association between clonal diversity and treatment resistance for at least some tumor types—notably ovarian (16) and esophageal (17). Basal-like TNBCs have previously been linked with shorter disease-free survival compared with nonbasal-like TNBCs and tend to be associated with higher clonal diversity (15). Furthermore, integration of genomic and drug response data from breast cancer PDX models generated in our laboratory shows that polygenomically engrafted tumors are more resistant to therapy than monogenomically engrafted tumors (Bruna and colleagues, manuscript in preparation). Clearly more work is still to be done, but it seems likely that the clonal composition of tumors will have future utility in predicting disease outcome and informing treatment choice.

Multiple groups have attempted to define clonal dynamics based on either lentiviral barcoding (cellular clones) or mutational clustering (genomic clones) by population and single cell-based computational approaches (5, 18). For example, Eirew and colleagues (5) reconstructed the genomic clonal dynamics of a panel of breast cancer PDX models using PyClone: a Bayesian clustering method for grouping somatic mutations. In each of the 15 cases examined, clonal diversity was reduced by xenotransplantation. This varied from extreme engraftment bias, selecting minor clones present in the sample of origin, to only moderate clonal selection. Remarkably, similar clonal dynamics were observed in parallel xenografts established from the same sample. In a separate study, Ding and colleagues (19) found that PDX models established from a basal-like breast cancer were more representative of the patient's metastatic lesion than the primary tumor. These observations suggest that deterministic mechanisms underline the clonal selection found on engraftment. Eirew and

Cassidy et al.

colleagues (1, 5) further observed variable clonal dynamics between PDXs established from different molecular subtypes, underscoring the need for better representation of tumor molecular subtypes. In summary, PDX models can, at least partially, recapitulate the complex clonal dynamics of human malignancies and engraftment biases may represent nonstochastic selection events, which define a PDX model rather than limit its utility. The variable tumor- and subtype-dependent engraftment rates (low for some tumor types) and frequencies in the population mean that PDX programs representative of intertumor heterogeneity may require large multicenter collaborative efforts (such as the EuroPDX consortium; ref. 10) and centralization of models.

Clonal evolution is a continuous process, and may be substantially altered by the selective pressures applied during chemo- and targeted therapy. Coupled with spatial heterogeneity within the tumor, this may result in the need for multiple-site repeat biopsies to decipher the clonal composition and dynamics of a tumor and inform treatment choice. Clearly these procedures will be highly invasive and may not be technically possible; it is with this in mind that many researchers have turned to liquid biopsies (14, 20). If our hypotheses are correct, modeling of clonal evolution in patient and matched PDX by computational approaches [such as PyClone (5) or CloneHD (18)] will allow unprecedented basic and translational research into clonal drivers of tumor progression and treatment response. Such studies will also be necessary to determine the longevity of PDX models of each cancer subtype, as it is unlikely that PDXs will remain patient-relevant and genomically stable in perpetuity.

It has been proposed that aberrant DNA methylation patterns in cancer can blur the lines between distinct phenotypic "attractor states" (21). Thus, it may be important to consider cellular clones as drivers of malignant progression, independent of their genetic background. To track these isogenic cellular clones, lentiviral tagging has been used to mark individual cells, and their progeny, in breast and colorectal PDX models (6, 7). In an elegant study by Kreso and colleagues (7), it was found that minor "type IV" subclones in colorectal PDXs were able to repopulate the tumor bulk after treatment with chemotherapy. These quiescent cell populations were genetically similar to their highly proliferative counterparts, and were later linked to the BMI1⁺ population thought to act as reserve stem cells of the intestinal and colonic crypts (22). As the cellular clones defined in this study were isogenic, their phenotype may have been driven by microenvironmental cues capable of modulating cellular transition between distinct gene expression patterns or epigenetic attractor states (23). Hence, the microenvironment's composition could profoundly alter both a cell's propensity to malignancy and the heterogeneity we hope to preserve in PDX models. Regardless, we should not underestimate the significance of even a partial translation of vastly heterogeneous human diseases into experimental model systems.

Limitations of the PDX Model: Tumor-Extrinsic Sources of Heterogeneity

Aside from clonal dynamics driven by intrinsic differences in a cell's genetic or epigenetic background, intratumor heterogeneity can be influenced by tumor-extrinsic factors in the non-cell-autonomous compartment (Fig. 1A and Table 1; ref. 23). Cellular interactions with the extracellular matrix (ECM) can alter gene expression programs, drive differentiation, and profoundly alter

cell behavior. As cancers develop, tight regulation of the ECM is lost and tissue architecture begins to degrade (8). A recent study by Wang and colleagues (24) provides direct evidence that ECM-dependent signaling confers dynamic switching between TGFBR3 (transforming growth factor β receptor 3)- and JUND (jun D proto-oncogene)-related expression signatures. ECM-driven oscillations between signaling pathways such as those described could have profound effects on propensity to malignancy. Furthermore, solid-state ECM interactions are necessary for cells to maintain stem cell properties and regulated ECM helps maintain the stem cell niche (25). As cancer is often associated with a blurring of the boundaries between stem and differentiated cells, it is possible that a loss of structured ECM is essential for the stability of multiple subdominant cellular clones within a tumor (21). In PDX models, Matrigel is often used to increase the engraftment efficiency; however, it is worth noting that this is a murine basement membrane extract and suitable synthetic human alternatives are available. The presence of growth factors in Matrigel may favor the engraftment of one cell type over another. Finally, as ECM structure is tissue specific (25), researchers should consider the use of orthotopic transplantations where possible.

The tumor microenvironment is further characterized by an influx of stromal cells. Infiltrating cancer-associated fibroblasts (CAF) can often confer resistance to cytotoxic and targeted therapies (9); however, recent studies confer on fibroblasts a degree of plasticity, with antitumor properties observed in some populations (26). Because of the high levels of CAF infiltrates seen in some tumor types, heterogeneity within their population would undoubtedly confer differential properties to the tumor bulk. We, and others, have found that human stromal cells are gradually replaced by murine equivalents upon engraftment in the mouse, suggesting that implanted human cancer cells retain the ability to recruit murine accessory cells to their niche. However, it should be noted that some differences exist between ligand repertoires of human and murine fibroblasts (27). Clearly stromal architecture and activity is mimicked in the murine host; however, it is currently unclear how this reflects human stroma with regards to supporting tumor growth and development.

It is with this in mind that many have begun to investigate the coengraftment of human mesenchymal stem cells (MSC) or CAF cell lines in PDXs (Fig. 1B and C). Here, care must be taken; the role of MSCs in tumor development is still controversial and may represent patient (or tissue)-specific differences (28). If non-patient-matched sources of stromal cells are implanted, heterogeneity between tumors derived from different patients could be lost. Moreover, fibroblast cell lines vary considerably in their ability to confer resistance to cytotoxic therapies through hepatocyte growth factor (HGF)/c-Met signaling (9, 27). Patient-derived fibroblasts can be isolated from tumor samples and expanded *in vitro*, thus, coengraftment of matched stromal components should be considered wherever possible. This would significantly increase the advantages these models already have in retaining the complex heterogeneity found in patient samples.

Considering the crucial role of the immune system in tumor progression, perhaps the most obvious disadvantage of PDX models is the necessity for severely immunodeficient host animals. Tumor cells are broadly thought to be antigenic: point mutations in coding exons in a developed tumor results in a large repertoire of neoantigens. Targeting of these neoantigens can lead to significant CD8⁺ cytotoxic T-cell infiltration and tumor cell death. However, most tumors eventually progress and evade

the immune system—often through the dominant inhibitory effects of suppressive pathways (so-called “immune checkpoints” such as CTLA-4/B7 and PD-1/PD-L1). This is supported by the prognostic value of the CD8⁺ to FOXP3⁺ (cytotoxic to regulatory T-cell, Treg) ratio in many solid tumors, and the recently reported clinical efficacy of a variety of checkpoint inhibitors (29, 30).

The proinflammatory microenvironment established by CD8⁺ T cells, M1 polarized tumor-associated macrophages (TAM), NK cells, and others can lead to the recruitment of numerous immune-suppressive components. TAMs and myeloid-derived suppressor cells have been implicated in resistance to antiangiogenic therapy. In addition, macrophage and CD4⁺ T-cell recruitment following intensive chemotherapy in breast cancer patients is associated with significantly reduced recurrence-free survival (30). Clearly, the coengraftment of immune components into PDX models (Fig. 1B) would facilitate both the study of novel therapies targeting tumor-immune interactions and allow for basic research into patient-specific cross-talk between tumor progression and immune surveillance. Clone or patient-specific differences in tolerization of dendritic cells, macrophage, and neutrophil polarity and Treg infiltration could provide multiple novel insights into tumor biology.

The most robust reconstitution of the human immune system in immunodeficient mice is seen when bone marrow–derived stem cells are coengrafted along with liver and thymus fragments (BMT model; ref. 31). However, the highly invasive multisite biopsy required renders this model impractical for patient-matched humanized PDX models (huPDXs). CD34 marks a population of hematopoietic stem and progenitor cells (HSPC) found in the blood and bone marrow. Mice implanted with CD34⁺ cells from umbilical cord blood show robust multilineage engraftment of human immune populations, though with reduced functionality relative to BMT mice. During chemotherapy, CD34⁺ cells leave the bone marrow and enter the circulation. If patients are given granulocyte-macrophage colony stimulating factor (GM-CSF; Leukine) to aid recovery from chemotherapy, numbers of CD34⁺ cells in the circulation have been known to exceed the bone marrow itself. Hence, CD34⁺ cells could potentially be harvested from patient's blood to reconstitute a functional, patient-matched, immune system in mouse models (huCD34 model). As such, an immune system would mature in the mouse and human immune cells would undergo central tolerance to mouse antigens during development. However, as they would not be exposed to patient antigens during this process, it is likely that they would mount a rapid non-self-response to any subsequently engrafted human tissue. One possible solution comes from a recent study by Cosgun and colleagues (32), showing immunodeficient adult mice carrying a mutation in the *Kit* receptor could support robust, uniform, and sustained engraftment of CD34⁺ cells. If patient-matched CD34⁺ HSPCs were engrafted in adult *Kit* mutant mice with an established PDX tumor, it is possible that a functional immune system could develop in the presence of both human and mouse antigens.

The alternative, to engraft fully mature human immune cells, is used in the huPBMC model. Here, mononuclear cells from the peripheral blood mononuclear cells (PBMC) are isolated and implanted in immunodeficient mice. PBMCs comprise around 75% CD4/CD8⁺ T cells with the remainder primarily containing B and NK cells. As such, these models are well suited to short-term experiments where lymphocyte function is of primary concern. Aside from the lifetime of circulating cells, the major limitation of

the huPBMC model is the rapid onset of graft versus host disease (GVHD) as engrafted cells mount an immune response against host murine tissue. The onset of GVHD can be delayed somewhat by the use of NOD *scid* gamma (NSG) strains lacking MHC I (33).

Undoubtedly, the reconstitution of a patient-matched immune system in PDX models would be extremely valuable in the development of novel oncologic drugs and, in particular, immunotherapeutics. The extent to which findings would be valid in the context of immune hyperactivation (against mouse in huPBMC or against tumor in huCD34) is unclear.

Conclusions and Future Prospects—Toward a huPDX Model

The PDX is arguably the most faithful model of cell-autonomous drivers of malignant progression. Sharing of expertise and resources through collaborative initiatives, such as the EuroPDX consortium, will lead to improved engraftment efficiencies and increase the breadth of tumor subtypes available in these model systems, contributing to a better representation of cancer heterogeneity in the laboratory (10).

Cooperation and competition between genetically and phenotypically distinct subclonal populations are thought to drive tumor growth, resistance to therapy, and recurrence. Although cell-autonomous sources of heterogeneity are clearly recapitulated in PDX models, current techniques fail to properly account for non-cell-autonomous factors. The microenvironment has long been known to play a significant role in tumor progression, but an incomplete understanding of stromal influences makes coengraftment of non-patient-matched cell compartments a risky strategy. In order to maintain PDXs as models of diversity of human malignancies, we must take care to engraft only patient-matched stromal components. CAFs have known roles in resistance to targeted therapy. It should, therefore, be a priority to ensure patient fibroblasts are maintained in the tumor xenograft.

Reconstituting a patient-matched immune system in PDX models is a significant challenge. Standard methods of engrafting either CD34⁺ HSPCs or mature circulating PBMCs will likely lead to inappropriate immune responses against human or murine tissues, respectively. In order to study long-term tumor-immune interactions, a reconstituted immune system must simultaneously maintain tolerance to its human donor and acquire tolerance to its new host. A solution could come from engrafting both mature and naïve cell types. Maturing cells in the thymus and bone marrow would acquire central tolerance to murine tissues, whereas mature cells in the periphery could act to suppress inappropriate responses against human antigens.

PDX models are capable of recapitulating the complexity of human malignancy remarkably well. These models have extraordinary utility in basic cancer research and beyond this have demonstrated clinical predictive power, allowing multiplexed screening of novel therapeutics *in vivo*. However, the aforementioned limitations of this model must be carefully considered when interpreting data. Although a significant amount of research is still needed, patient-matched huPDX models with coengrafted stromal and immune components would offer unprecedented opportunity to study tumor biology and would be invaluable models in oncologic drug development. In the future, huPDX models could allow researchers and clinicians to both predict and explain tumor response to novel targeted therapies.

Cassidy et al.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Establishment of our Patient-Derived Tumour Xenograft models would not be possible without the dedicated work of the Cambridge Breast Unit, the Cambridge Institute Biological Resources Unit, and our collaborators in

EuroPDX consortium (<http://www.EuroPDX.eu>). J.W. Cassidy is grateful to O.B. Harris for critical review of the first draft of this manuscript. In addition, the authors are particularly grateful to the patients who donated tissues to further our research and Cancer Research UK for supporting all authors.

Received March 16, 2015; revised April 27, 2015; accepted April 27, 2015; published OnlineFirst July 15, 2015.

References

1. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;486:346–52.
2. Huang M, Shen A, Ding J, Geng M. Molecularly targeted cancer therapy: some lessons from the past decade. *Trends Pharmacol Sci* 2014;35:41–50.
3. Shaw AT, Kim D-W, Nakagawa K, Seto T, Crinó L, Ahn M-J, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med* 2013;368:2385–94.
4. Diaz LA, Williams RT, Wu J, Kinzler I, Hecht JR, Berlin J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 2012;486:537–40.
5. Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* 2014;518:422–6.
6. Nguyen LV, Cox CL, Eirew P, Knapp DJHF, Pellacani D, Kannan N, et al. DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts. *Nat Commun* 2014;5:5871.
7. Kreso A, O'Brien CA, van Galen P, Gan OI, Notta F, Brown AMK, et al. Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science* 2013;339:543–8.
8. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med* 2013;19:1423–37.
9. Straussman R, Morikawa T, Shee K, Barzily-Rokni M, Qian ZR, Du J, et al. Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. *Nature* 2012;487:500–4.
10. Hidalgo M, Amant F, Biankin AV, Budinska E, Byrne AT, Caldas C, et al. Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discov* 2014;4:998–1013.
11. Whittle JR, Lewis MT, Lindeman GJ, Visvader JE. Patient-derived xenograft models of breast cancer and their predictive power. *Breast Cancer Res* 2015;17. doi: 10.1186/s13058-015-0523-1.
12. Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* 2011;146:633–44.
13. Cleary AS, Leonard TL, Gestl SA, Gunther EJ. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature* 2014;508:113–7.
14. Aparicio S, Caldas C. The implications of clonal genome evolution for cancer medicine. *N Engl J Med* 2013;368:842–51.
15. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 2012;486:395–9.
16. Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, et al. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J Pathol* 2013;231:21–34.
17. Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, Sanchez CA, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat Genet* 2006;38:468–73.
18. Fischer A, Vázquez-García I, Illingworth CJR, Mustonen V. High-definition reconstruction of clonal composition in cancer. *Cell Rep* 2014;7:1740–52.
19. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* 2010;464:999–1005.
20. Murtaza M, Dawson S-J, Tsui DWY, Gale D, Forshaw T, Piskorz AM, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* 2013;497:108–12.
21. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* 2012;12:323–34.
22. Kreso A, van Galen P, Pedley NM, Lima-Fernandes E, Frelin C, Davis T, et al. Self-renewal as a therapeutic target in human colorectal cancer. *Nat Med* 2014;20:29–36.
23. Michor F, Weaver VM. Understanding tissue context influences on intra-tumour heterogeneity. *Nat Cell Biol* 2014;16:301–2.
24. Wang C-C, Bajlikar SS, Jamal L, Atkins KA, Janes KA. A time- and matrix-dependent TGFBR3-JUND-KRT5 regulatory circuit in single breast epithelial cells and basal-like premalignancies. *Nat Cell Biol* 2014;16:345–56.
25. Cassidy JW. Nanotechnology in the regeneration of complex tissues. *Bone Tissue Regen Insights* 2014;5:25–35.
26. Augsten M. Cancer-associated fibroblasts as another polarized cell type of the tumor microenvironment. *Front Oncol* 2014;4:62.
27. Argent R, Kumari R, Clarke P, Onion D, Lobo D, Grabowska A, et al. Restoration of paracrine signalling within the tumour microenvironment increases tumour growth and activation of c-Met [abstract]. In: Proceedings of the 10th NCRI Cancer Conference Poster Session C. Nottingham: NCRI; 2014. Abstract nr. LB186.
28. Yagi H, Kitagawa Y. The role of mesenchymal stem cells in cancer development. *Front Genet* 2013;4:261.
29. Gubin MM, Zhang X, Schuster H, Caron E, Ward JP, Noguchi T, et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* 2014;515:577–81.
30. Schmidt M, Böhm D, Von Törne C, Steiner E, Puhl H, Pilch H, et al. The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res* 2008;68:5405–13.
31. Shultz LD, Ishikawa F, Greiner DL. Humanized mice in translational biomedical research. *Nat Rev Immunol* 2007;7:118–30.
32. Cosgun KN, Rahmig S, Mende N, Reinke S, Hauber I, Schäfer C, et al. Kit regulates HSC engraftment across the human-mouse species barrier. *Cell Stem Cell* 2014;15:227–38.
33. King M, Pearson T, Shultz LD, Leif J, Bottino R, Trucco M, et al. A new Hu-PBL model for the study of human islet alloreactivity based on NOD-scid mice bearing a targeted mutation in the IL-2 receptor gamma chain gene. *Clin Immunol* 2008;126:303–14.

References

1. Heppner GH. Tumor heterogeneity. *Cancer Res.* 1984;44(6):2259–65.
2. Dexter D, Spremulli E, Fligiel Z. Heterogeneity of cancer cells from a single human colon carcinoma. *Am J Med.* 1978 Aug;71(6):A40.
3. Danielson KG, Anderson LW, Hosick HL. Selection and Characterization in Culture of Mammary Tumor Cells with Distinctive Growth Properties in Vivo Selection and Characterization in Culture of Mammary Tumor Cells with Distinctive Growth Properties in Vivo1. *Cancer Res.* 1980;(JULY 1980):1812–9.
4. Barranco SC, Ho DHW, Drewinko B, Romsdahl MM, Humphrey RM. Differential Sensitivities of Human Melanoma Cells Grown in Vitro to Arabinosylcytosine Differential Sensitivities of Human Melanoma Cells Grown in Vitro to Arabinosylcytosine I. *Chart.* 1972;(DECEMBER):2733–6.
5. Chang K, Creighton CJ, Davis C, Donehower L, Drummond J, Wheeler D, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* 2013;45(10):1113–20.
6. Cole MP, Jones CTA, Todd IDH. A New Anti-oestrogenic Agent in Late Breast Cancer: An Early Clinical Appraisal of ICI46474. *Br J Cancer.* 1971 Jun;25(2):270–5.
7. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012 Jun;486(7403):346–52.
8. Chia SK, Bramwell VH, Tu D, Shepherd LE, Jiang S, Vickery T, et al. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Clin Cancer Res [Internet].* 2012 Aug 15 [cited 2018 Sep 18];18(16):4465–72. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22711706>
9. Liu MC, Pitcher BN, Mardis ER, Davies SR, Friedman PN, Snider JE, et al. PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: correlative analysis of C9741 (Alliance). *npj Breast Cancer [Internet].* 2016 Dec 6 [cited 2018 Sep 18];2(1):15023. Available from: <http://www.nature.com/articles/npjbcancer201523>
10. Ali HR, Rueda OM, Chin S-F, Curtis C, Dunning MJ, Aparicio S, et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. *Genome Biol.* 2014;15(8):431.

11. Cassidy JW, Bruna A. Chapter 4 - Tumor Heterogeneity BT - Patient Derived Tumor Xenograft Models. In Academic Press; 2017. p. 37–55. Available from: <https://www.sciencedirect.com/science/article/pii/B9780128040102000047>
12. Nowell PC. The clonal evolution of tumor cell populations. *Sci*. 1976 Oct;194(4260):23–8.
13. Fidler I, Kripke M. Metastasis results from preexisting variant cells within a malignant tumor. *Science* (80-). 1977;197(4306):893–5.
14. GAVOSTO F, PILERI A, GABUTTI V, MASERA P. Non-self-maintaining Kinetics of Proliferating Blasts in Human Acute Leukaemia. *Nature*. 1967 Oct;216(5111):188–9.
15. Clarkson B, Fried J, Strife A, Sakai Y, Ota K, Ohkita T. Studies of cellular proliferation in human leukemia. III. Behavior of leukemic cells in three adults with acute leukemia given continuous infusions of 3H-thymidine for 8 or 10 days. *Cancer*. 1970 Jun;25(6):1237–60.
16. Clarkson B, Ohkita T, Ota K, Fried J. Studies of Cellular Proliferation in Human Leukemia. I. Estimation of Growth Rates of Leukemic and Normal Hematopoietic Cells in Two Adults with Acute Leukemia Given Single Injections of Tritiated Thymidine *. *J Clin Invest*. 1967 Apr;46(4):506–29.
17. Southam C, Brunschwig A, Dizon Q. Autologous and homologous transplantation of human cancer. In: Brennan MJ, Simpson WL, editors. *Biological interactions in normal and neoplastic growth: a contribution to the tumor-host problem*. Boston: Little, Brown; 1962. p. 723–738.
18. Pierce G, Dixon F, Verney E. Teratocarcinogenic and tissue-forming potentials of the cell types comprising neoplastic embryoid bodies. *Lab Invest*. 1960;9(1):583–602.
19. Baylin SB, Weisburger WR, Eggleston JC, Mendelsohn G, Beaven MA, Abeloff MD, et al. Variable Content of Histaminase, L-Dopa Decarboxylase and Calcitonin in Small-Cell Carcinoma of the Lung. *N Engl J Med*. 1978 Jul;299(3):105–10.
20. Bennett DC, Peachey LA, Durbin H, Rudland PS. A possible mammary stem cell line. *Cell*. 1978 Sep;15(1):283–98.
21. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic Alterations during Colorectal-Tumor Development. *N Engl J Med*. 1988 Sep;319(9):525–32.
22. Auman JT, McLeod HL. Colorectal Cancer Cell Lines Lack the Molecular Heterogeneity of Clinical Colorectal Tumors. *Clin Colorectal Cancer*. 2015 Sep;9(1):40–7.

23. Cassidy JW, Caldas C, Bruna A. Maintaining Heterogeneity in Patient Derived Tumour Xenografts. *Cancer Res.* 2015;75(15):1–6.
24. Kreso A, Dick JE. Evolution of the cancer stem cell model. *Cell Stem Cell.* 2014;14(3):275–91.
25. Sansom OJ, Reed KR, Hayes AJ, Ireland H, Brinkmann H, Newton IP, et al. Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration. 2004;1385–90.
26. Barker N, Ridgway R a, van Es JH, van de Wetering M, Begthel H, van den Born M, et al. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature.* 2009 Jan;457(7229):608–11.
27. Cairns J. Mutation selection and the natural history of cancer. *Nature.* 1975 May;255(5505):197–200.
28. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. *Cell.* 2012;149(5):994–1007.
29. Tabassum DP, Polyak K. Tumorigenesis: it takes a village. *Nat Rev Cancer.* 2015;(July):1–11.
30. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Vol. 144, *Cell.* 2011. p. 646–74.
31. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* [Internet]. 2000 Jan 7 [cited 2018 Sep 18];100(1):57–70. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10647931>
32. Andor N, Maley CC, Ji HP. Genomic Instability in Cancer: Teetering on the Limit of Tolerance. *Cancer Res* [Internet]. 2017 May 1 [cited 2018 Sep 18];77(9):2179–85. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28432052>
33. Sniegowski PD, Gerrish PJ, Johnson T, Shaver A. The evolution of mutation rates: separating causes from consequences. *BioEssays* [Internet]. 2000 Nov 10 [cited 2018 Sep 18];22(12):1057–66. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11084621>
34. Nowak M, Schuster P. Error thresholds of replication in finite populations mutation frequencies and the onset of Muller’s ratchet. *J Theor Biol* [Internet]. 1989 Apr 20 [cited 2018 Sep 18];137(4):375–95. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2626057>
35. Bagnoli F, C MB-IJ of MP, 1998 undefined. Eigen’s error threshold and mutational meltdown in a quasispecies model. *World Sci* [Internet]. [cited 2018 Sep 18]; Available

- from: <https://www.worldscientific.com/doi/pdf/10.1142/S0129183198000935>
36. Herr AJ, Ogawa M, Lawrence NA, Williams LN, Eggington JM, Singh M, et al. Mutator Suppression and Escape from Replication Error–Induced Extinction in Yeast. Copenhaver GP, editor. PLoS Genet [Internet]. 2011 Oct 6 [cited 2018 Sep 18];7(10):e1002282. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22022273>
 37. Crotty S, Cameron CE, Andino R. RNA virus error catastrophe: direct molecular test by using ribavirin. Proc Natl Acad Sci U S A [Internet]. 2001 Jun 5 [cited 2018 Sep 18];98(12):6895–900. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11371613>
 38. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013;45(10):1127–33.
 39. Vesely M, of RS-A of the NYA, 2013 undefined. Cancer immunoediting: antigens, mechanisms, and implications to cancer immunotherapy. Wiley Online Libr [Internet]. [cited 2018 Sep 18]; Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/nyas.12105>
 40. Santarpia M, medicine NK-C biology &, 2015 undefined. Tumor immune microenvironment characterization and response to anti-PD-1 therapy. ncbi.nlm.nih.gov [Internet]. [cited 2018 Sep 18]; Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4493379/>
 41. Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. Genome Biol [Internet]. 2015 Dec 31 [cited 2018 Sep 18];16(1):64. Available from: <http://genomebiology.com/2015/16/1/64>
 42. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med [Internet]. 2010 Aug 19 [cited 2018 Sep 18];363(8):711–23. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20525992>
 43. Gubin MM, Zhang X, Schuster H, Caron E, Ward JP, Noguchi T, et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature. 2014 Nov;515(7528):577–81.
 44. Eigen M. Error catastrophe and antiviral strategy. Proc Natl Acad Sci U S A [Internet]. 2002

- Oct 15 [cited 2018 Sep 18];99(21):13374–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12370416>
45. Jiang X, Mu B, Huang Z, Zhang M, Wang X, Tao S. Impacts of mutation effects and population size on mutation rate in asexual populations: a simulation study. *BMC Evol Biol* [Internet]. 2010 Sep 30 [cited 2018 Sep 18];10:298. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20920286>
 46. Beckman RA. Mutator mutations enhance tumorigenic efficiency across fitness landscapes. *PLoS One* [Internet]. 2009 Jun 10 [cited 2018 Sep 18];4(6):e5860. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19517009>
 47. Schrödinger E. What Is Life? the physical aspect of the living cell and mind [Internet]. 1944 [cited 2018 Sep 18]. Available from: <http://www.spaz.org/~jake/pix/schrodinger.pdf>
 48. Biebricher CK, Eigen M. The error threshold. *Virus Res* [Internet]. 2005 Feb [cited 2018 Sep 18];107(2):117–27. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15649558>
 49. Crotty S, ... CC-P of the, 2001 undefined. RNA virus error catastrophe: direct molecular test by using ribavirin. *Natl Acad Sci* [Internet]. [cited 2018 Sep 18]; Available from: <http://www.pnas.org/content/98/12/6895.short>
 50. Zeyl C, Mizesko M, Evolution JDV-, 2001 undefined. Mutational meltdown in laboratory yeast populations. *Wiley Online Libr* [Internet]. [cited 2018 Sep 18]; Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0014-3820.2001.tb00608.x>
 51. Lynch M, Bürger R, Butcher D, Heredity WG-J of, 1993 undefined. The mutational meltdown in asexual populations. *academic.oup.com* [Internet]. [cited 2018 Sep 18]; Available from: <https://academic.oup.com/jhered/article-abstract/84/5/339/2186429>
 52. Robles AI, Harris CC. Clinical outcomes and correlates of TP53 mutations and cancer. *Cold Spring Harb Perspect Biol* [Internet]. 2010 Mar [cited 2018 Sep 18];2(3):a001016. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20300207>
 53. Gambino V, De Michele G, Venezia O, Migliaccio P, Dall'Olio V, Bernard L, et al. Oxidative stress activates a specific p53 transcriptional response that regulates cellular senescence and aging. *Aging Cell* [Internet]. 2013 Jun [cited 2018 Sep 18];12(3):435–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23448364>
 54. Degtyareva NP, Heyburn L, Sterling J, Resnick MA, Gordenin DA, Doetsch PW. Oxidative stress-induced mutagenesis in single-strand DNA occurs primarily at cytosines and is

- DNA polymerase zeta-dependent only for adenines and guanines. *Nucleic Acids Res* [Internet]. 2013 Oct [cited 2018 Sep 18];41(19):8995–9005. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23925127>
55. Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledó JI, Howe DK, et al. A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proc Natl Acad Sci U S A* [Internet]. 2009 Sep 22 [cited 2018 Sep 18];106(38):16310–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19805298>
 56. Mroz EA, Tward AD, Tward AM, Hammon RJ, Ren Y, Rocco JW. Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. *PLoS Med* [Internet]. 2015 Feb [cited 2018 Sep 18];12(2):e1001786. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25668320>
 57. Chen G, Mulla WA, Kucharavy A, Tsai H-J, Rubinstein B, Conkright J, et al. Targeting the adaptability of heterogeneous aneuploids. *Cell* [Internet]. 2015 Feb 12 [cited 2018 Sep 18];160(4):771–84. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25679766>
 58. Chen G, Bradford WD, Seidel CW, Li R. Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy. *Nature* [Internet]. 2012 Jan 29 [cited 2018 Sep 18];482(7384):246–50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22286062>
 59. Birnbak NJ, Eklund AC, Li Q, McClelland SE, Endesfelder D, Tan P, et al. Paradoxical relationship between chromosomal instability and survival outcome in cancer. *Cancer Res* [Internet]. 2011 May 15 [cited 2018 Sep 18];71(10):3447–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21270108>
 60. Roylance R, Endesfelder D, Gorman P, Burrell RA, Sander J, Tomlinson I, et al. Relationship of extreme chromosomal instability with long-term survival in a retrospective analysis of primary breast cancer. *Cancer Epidemiol Biomarkers Prev* [Internet]. 2011 Oct [cited 2018 Sep 18];20(10):2183–94. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21784954>
 61. Ke K, Cheng J, Hunt AJ. The distribution of polar ejection forces determines the amplitude of chromosome directional instability. *Curr Biol* [Internet]. 2009 May 26 [cited 2018 Sep 18];19(10):807–15. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19446456>
 62. Fehrmann RSN, Karjalainen JM, Krajewska M, Westra H-J, Maloney D, Simeonov A, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet*

- [Internet]. 2015 Feb 12 [cited 2018 Sep 18];47(2):115–25. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25581432>
63. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nat Rev Genet* [Internet]. 2009 Aug [cited 2018 Sep 18];10(8):551–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19597530>
 64. Nelson WG, Kastan MB. DNA strand breaks: the DNA template alterations that trigger p53-dependent DNA damage response pathways. *Mol Cell Biol* [Internet]. 1994 Mar [cited 2018 Sep 18];14(3):1815–23. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8114714>
 65. Holst F, Stahl PR, Ruiz C, Hellwinkel O, Jehan Z, Wendland M, et al. Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer. *Nat Genet*. 2007 May;39(5):655–60.
 66. Kallioniemi OP, Kallioniemi a, Kurisu W, Thor a, Chen LC, Smith HS, et al. ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. *Proc Natl Acad Sci U S A*. 1992;89(12):5321–5.
 67. Schnitt SJ. Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. *Mod Pathol*. 2010 May;23(S2):S60–4.
 68. Ortmann CA, Kent DG, Nangalia J, Silber Y, Wedge DC, Grinfeld J, et al. Effect of Mutation Order on Myeloproliferative Neoplasms. *N Engl J Med*. 2015 Feb;372(7):601–12.
 69. Cancer T, Atlas G. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012 Jul;487(7407):330–7.
 70. Barnetson R, Jass J, Tse R, Eckstein R, Robinson B, Schnitzler M. Mutations associated with microsatellite unstable colorectal carcinomas exhibit widespread intratumoral heterogeneity. *Genes, Chromosom Cancer*. 2000 Oct;29(2):130–6.
 71. Koren S, Reavie L, Couto JP, De Silva D, Stadler MB, Roloff T, et al. PIK3CAH1047R induces multipotency and multi-lineage mammary tumours. *Nature*. 2015 Sep;525(7567):114–8.
 72. Van Keymeulen A, Lee MY, Ousset M, Brohee S, Rorive S, Giraddi RR, et al. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature*. 2015 Aug;advance on.
 73. Pereira B, Chin S-F, Rueda OM, Volland H-KM, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat Commun*. 2016 May;7.

74. Bhat-Nakshatri P, Goswami CP, Badve S, Magnani L, Lupien M, Nakshatri H. Molecular Insights of Pathways Resulting from Two Common PIK3CA Mutations in Breast Cancer. *Cancer Res* [Internet]. 2016 Jul 1 [cited 2018 Sep 24];76(13):3989–4001. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27197157>
75. Nakshatri H, Goswami C, Badve S, Magnani L, Lupien M, Bhat-Nakshatri P. Abstract P3-05-15: Divergent activation of AKT1 and AKT2 isoforms downstream of PI3K mutation impacts response of breast cancer cells to estradiol and PI3K inhibitors. *Cancer Res* [Internet]. 2015 May 1 [cited 2018 Sep 24];75(9 Supplement):P3-05-15-P3-05-15. Available from: <http://cancerres.aacrjournals.org/lookup/doi/10.1158/1538-7445.SABCS14-P3-05-15>
76. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin A V, et al. Signatures of mutational processes in human cancer. *Nature*. 2013 Aug;500(7463):415–21.
77. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Sci* . 2015 May;348(6237):880–6.
78. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012 Jun;486(7403):395–9.
79. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet*. 2014 Mar;46(3):225–33.
80. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Sci* . 2014 Oct;346(6206):251–6.
81. Caravagna G, Giarratano Y, Ramazzotti D, Tomlinson I, Graham TA, Sanguinetti G, et al. Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nat Methods* [Internet]. 2018 Sep [cited 2018 Sep 27];15(9):707–14. Available from: <http://www.nature.com/articles/s41592-018-0108-x>
82. Cleary AS, Leonard TL, Gestl S a, Gunther EJ. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature*. 2014 Apr;508(7494):113–7.
83. Kreso a., O'Brien C a., van Galen P, Gan OI, Notta F, Brown a. MK, et al. Variable Clonal

- Repopulation Dynamics Influence Chemotherapy Response in Colorectal Cancer. *Science* (80-). 2012;339(6119):543–8.
84. Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat Rev Cancer* [Internet]. 2013 Jul 13 [cited 2018 Sep 24];13(7):497–510. Available from: <http://www.nature.com/articles/nrc3486>
 85. Polak P, Karlić R, Koren A, Thurman R, Sandstrom R, Lawrence MS, et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* [Internet]. 2015 Feb 19 [cited 2018 Sep 24];518(7539):360–4. Available from: <http://www.nature.com/articles/nature14221>
 86. Sharma S V., Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. *Cell* [Internet]. 2010 Apr 2 [cited 2018 Sep 24];141(1):69–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20371346>
 87. Wu Ct C -t., Morris JR. Genes, genetics, and epigenetics: a correspondence. *Science* [Internet]. 2001 Aug 10 [cited 2018 Sep 26];293(5532):1103–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11498582>
 88. Baylin SB, Jones PA. A decade of exploring the cancer epigenome — biological and translational implications. *Nat Rev Cancer* [Internet]. 2011 Oct 1 [cited 2018 Sep 26];11(10):726–34. Available from: <http://www.nature.com/articles/nrc3130>
 89. Ehrlich M, Woods CB, Yu MC, Dubeau L, Yang F, Campan M, et al. Quantitative analysis of associations between DNA hypermethylation, hypomethylation and DNMT RNA levels in ovarian tumors. *Oncogene* [Internet]. 2006 Apr 13 [cited 2018 Sep 26];25(18):2636–45. Available from: <http://www.nature.com/articles/1209145>
 90. De S, Shaknovich R, Riester M, Elemento O, Geng H, Kormaksson M, et al. Aberration in DNA Methylation in B-Cell Lymphomas Has a Complex Origin and Increases with Disease Severity. Grimes HL, editor. *PLoS Genet* [Internet]. 2013 Jan 10 [cited 2018 Sep 26];9(1):e1003137. Available from: <http://dx.plos.org/10.1371/journal.pgen.1003137>
 91. Domcke S, Bardet AF, Adrian Ginno P, Hartl D, Burger L, Schübeler D. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* [Internet]. 2015 Dec 16 [cited 2018 Sep 26];528(7583):575–9. Available from: <http://www.nature.com/articles/nature16462>
 92. Assenov Y, Brocks D, Gerhäuser C. Intratumor heterogeneity in epigenetic patterns. *Semin*

- Cancer Biol [Internet]. 2018 Aug [cited 2018 Sep 24];51:12–21. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1044579X17302262>
93. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in epigenetic domains across cancer types. Nat Genet [Internet]. 2011 Aug 26 [cited 2018 Sep 26];43(8):768–75. Available from: <http://www.nature.com/articles/ng.865>
94. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer [Internet]. 2012 May 19 [cited 2018 Sep 14];12(5):323–34. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22513401>
95. You JS, Jones PA. Cancer Genetics and Epigenetics: Two Sides of the Same Coin? Cancer Cell [Internet]. 2012 Jul 10 [cited 2018 Sep 26];22(1):9–20. Available from: <https://www.sciencedirect.com/science/article/pii/S1535610812002577>
96. Lundstrom K. Personalized Medicine and Epigenetic Drug Development. In: Personalized Epigenetics [Internet]. Elsevier; 2015 [cited 2018 Sep 26]. p. 369–86. Available from: <https://linkinghub.elsevier.com/retrieve/pii/B9780124201354000139>
97. Stone A, Valdés-Mora F, Gee JMW, Farrow L, McClelland RA, Fiegl H, et al. Tamoxifen-induced epigenetic silencing of oestrogen-regulated genes in anti-hormone resistant breast cancer. PLoS One [Internet]. 2012 [cited 2018 Aug 22];7(7):e40466. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22808167>
98. Magnani L, Louloui A, Zwart W. Histone Posttranslational Modifications in Breast Cancer and Their Use in Clinical Diagnosis and Prognosis. Epigenetic Biomarkers and Diagnostics [Internet]. 2016 Jan 1 [cited 2018 Sep 24];467–77. Available from: <https://www.sciencedirect.com/science/article/pii/B9780128018996000231?via%3Dihub>
99. Dvinge H, Git A, Gräf S, Salmon-Divon M, Curtis C, Sottoriva A, et al. The shaping and functional consequences of the microRNA landscape in breast cancer. Nature [Internet]. 2013 May 5 [cited 2018 Sep 27];497(7449):378–82. Available from: <http://www.nature.com/articles/nature12108>
100. Le Quesne J, Caldas C. Micro-RNAs and breast cancer. Mol Oncol [Internet]. 2010 Jun [cited 2018 Sep 27];4(3):230–41. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20537965>
101. Buffa FM, Camps C, Winchester L, Snell CE, Gee HE, Sheldon H, et al. microRNA-Associated

- Progression Pathways and Potential Therapeutic Targets Identified by Integrated mRNA and microRNA Expression Profiling in Breast Cancer. *Cancer Res* [Internet]. 2011 Sep 1 [cited 2018 Sep 27];71(17):5635–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21737487>
102. Jurmeister S, Baumann M, Balwierz a., Keklikoglou I, Ward a., Uhlmann S, et al. MicroRNA-200c Represses Migration and Invasion of Breast Cancer Cells by Targeting Actin-Regulatory Proteins FHOD1 and PPM1F. *Mol Cell Biol*. 2012;32:633–51.
 103. Ottaviani S, Stebbing J, Frampton AE, Zagorac S, Krell J, de Giorgio A, et al. TGF- β induces miR-100 and miR-125b but blocks let-7a through LIN28B controlling PDAC progression. *Nat Commun* [Internet]. 2018 Dec 10 [cited 2018 Sep 27];9(1):1845. Available from: <http://www.nature.com/articles/s41467-018-03962-x>
 104. Ward A, Balwierz A, Zhang JD, Küblbeck M, Pawitan Y, Hielscher T, et al. Re-expression of microRNA-375 reverses both tamoxifen resistance and accompanying EMT-like properties in breast cancer. *Oncogene* [Internet]. 2013 Feb 16 [cited 2018 Aug 24];32(9):1173–82. Available from: <http://www.nature.com/articles/onc2012128>
 105. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* [Internet]. 2011 Jan 1 [cited 2018 Sep 24];39(Database):D945–50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20952405>
 106. Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng T, et al. *ARID1A* Mutations in Endometriosis-Associated Ovarian Carcinomas. *N Engl J Med* [Internet]. 2010 Oct 14 [cited 2018 Sep 24];363(16):1532–43. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20942669>
 107. Jones S, Wang T-L, Shih I-M, Mao T-L, Nakayama K, Roden R, et al. Frequent Mutations of Chromatin Remodeling Gene *ARID1A* in Ovarian Clear Cell Carcinoma. *Science* (80-) [Internet]. 2010 Oct 8 [cited 2018 Sep 24];330(6001):228–31. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20826764>
 108. Dawson MA. The cancer epigenome: Concepts, challenges, and therapeutic opportunities. *Science* [Internet]. 2017 Mar 17 [cited 2018 Sep 25];355(6330):1147–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28302822>
 109. Perera D, Poulos RC, Shah A, Beck D, Pimanda JE, Wong JWH. Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. *Nature* [Internet].

- 2016 Apr 14 [cited 2018 Sep 25];532(7598):259–63. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27075100>
110. Katainen R, Dave K, Pitkänen E, Palin K, Kivioja T, Välimäki N, et al. CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat Genet* [Internet]. 2015 Jul 8 [cited 2018 Sep 25];47(7):818–21. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26053496>
 111. Huang F, Hodis E, Xu M, Kryukov G, ... LC-, 2013 undefined. Highly recurrent TERT promoter mutations in human melanoma. *science.sciencemag.org* [Internet]. [cited 2018 Sep 25]; Available from: <http://science.sciencemag.org/content/339/6122/957.short>
 112. Horn S, Figl A, Rachakonda P, ... CF-, 2013 undefined. TERT promoter mutations in familial and sporadic melanoma. *science.sciencemag.org* [Internet]. [cited 2018 Sep 25]; Available from: <http://science.sciencemag.org/content/339/6122/959.short>
 113. Shimoi T, Yoshida M, Kitamura Y, Yoshino T, Kawachi A, Shimomura A, et al. TERT promoter hotspot mutations in breast cancer. *Breast Cancer* [Internet]. 2018 May 8 [cited 2018 Sep 27];25(3):292–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29222734>
 114. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*. 2013 Sep;501(7467):346–54.
 115. Pagès F, Galon J, Dieu-Nosjean M-C, Tartour E, Sautès-Fridman C, Fridman W-H. Immune infiltration in human tumors: a prognostic factor that should not be ignored. *Oncogene*. 2010;29(8):1093–102.
 116. Hadrup S, Donia M, Thor Straten P. Effector CD4 and CD8 T cells and their role in the tumor microenvironment. *Cancer Microenviron* [Internet]. 2013 Aug [cited 2018 Sep 18];6(2):123–33. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23242673>
 117. Mei Z, Liu Y, Liu C, Cui a, Liang Z, Wang G, et al. Tumour-infiltrating inflammation and prognosis in colorectal cancer: systematic review and meta-analysis. *Br J Cancer*. 2014;110(6):1595–605.
 118. de Kruijf E, Engels C, van de Water W, Bastiaannet E, Smit VHBM, van de Velde CH, et al. Tumor immune subtypes distinguish tumor subclasses with clinical implications in breast cancer patients. *Breast Cancer Res Treat*. 2013;142(2):355–64.
 119. Ali HR, Glont SS-E, Blows FM, Provenzano E, Dawson S-JS, Liu B, et al. PD-L1 protein expression in breast cancer is rare, enriched in basal-like tumours and associated with infiltrating lymphocytes. *Ann Oncol*. 2015;(April):1488–93.

120. Gibson J. Anti-PD-L1 for metastatic triple-negative breast cancer. *Lancet Oncol*. 2015;16(6):e264.
121. Waddington CH. The strategy of the genes. A discussion of some aspects of theoretical biology. Taylor & Francis; 1957. 262 p.
122. Cassidy JW, Roberts JN, Smith C-A, Robertson M, White K, Biggs MJ, et al. Osteogenic lineage restriction by osteoprogenitors cultured on nanometric grooved surfaces: The role of focal adhesion maturation(). *Acta Biomater* [Internet]. 2014 Feb 23;10(2):651–60. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3907683/>
123. Mohn F, Schübeler D. Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends Genet*. 2009;25(3):129–36.
124. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer*. 2012 May;12(5):323–34.
125. Quail DF, Joyce J a. Microenvironmental regulation of tumor progression and metastasis. *Nat Med*. 2013 Nov;19(11):1423–37.
126. Cassidy JW. Nanotechnology in the Regeneration of Complex Tissues. *Bone tissue Regen insights* [Internet]. 2014 Nov 12;5:25–35. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4471123/>
127. Wang C-C, Bajikar SS, Jamal L, Atkins K a, Janes K a. A time- and matrix-dependent TGFB3-JUND-KRT5 regulatory circuit in single breast epithelial cells and basal-like premalignancies. *Nat Cell Biol*. 2014 Apr;16(4):345–56.
128. Hutchinson L, Kirk R. High drug attrition rates--where are we going wrong? *Nat Rev Clin Oncol*. 2011;8(4):189–90.
129. Moreno L, Pearson ADJ. How can attrition rates be reduced in cancer drug discovery? *Expert Opin Drug Discov*. 2013 Feb;8(4):363–8.
130. Aparicio S, Caldas C. The implications of clonal genome evolution for cancer medicine. *N Engl J Med*. 2013 Feb;368(9):842–51.
131. Weigelt B, Reis-Filho JS. Histological and molecular types of breast cancer: is there a unifying taxonomy? *Nat Rev Clin Oncol*. 2009 Dec;6(12):718–30.
132. Engels CC, Fontein DBY, Kuppen PJK, de Kruijf EM, Smit VTHBM, Nortier JWR, et al. Immunological subtypes in breast cancer are prognostic for invasive ductal but not for invasive lobular breast carcinoma. *Br J Cancer*. 2014 Jul;111(3):532–8.
133. Lal N. An immunogenomic stratification of colorectal cancer: implications for

- development of targeted immunotherapy. *Oncoimmunology*. 2015;(September).
134. Diaz L a, Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature*. 2012 Jun;486(7404):537–40.
 135. Bhang HC, Ruddy D a, Krishnamurthy Radhakrishna V, Caushi JX, Zhao R, Hims MM, et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med*. 2015;21(5).
 136. Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, et al. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J Pathol*. 2013 Sep;231(1):21–34.
 137. Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, Sanchez C a, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat Genet*. 2006 Apr;38(4):468–73.
 138. Misale S, Di Nicolantonio F, Sartore-Bianchi a, Siena S, Bardelli a. Resistance to Anti-EGFR Therapy in Colorectal Cancer: From Heterogeneity to Convergent Evolution. *Cancer Discov*. 2014;4(11):1269–80.
 139. Lord CJ, Ashworth A. Mechanisms of resistance to therapies targeting BRCA-mutant cancers. *Nat Med*. 2013 Nov;19(11):1381–8.
 140. Xu G, Chapman JR, Brandsma I, Yuan J, Mistrik M, Bouwman P, et al. REV7 counteracts DNA double-strand break resection and affects PARP inhibition. *Nature*. 2015;521(7553):541–4.
 141. Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med* [Internet]. 2016 Jan 30 [cited 2018 Sep 18];22(1):105–13. Available from: <http://www.nature.com/articles/nm.3984>
 142. Kostadinov R, Maley CC, Kuhner MK. Bulk Genotyping of Biopsies Can Create Spurious Evidence for Heterogeneity in Mutation Content. Wang E, editor. *PLOS Comput Biol* [Internet]. 2016 Apr 22 [cited 2018 Sep 18];12(4):e1004413. Available from: <http://dx.plos.org/10.1371/journal.pcbi.1004413>
 143. Jiang L, Chen H, Pinello L, Yuan G-C. GiniClust: detecting rare cell types from single-cell gene expression data with Gini index. *Genome Biol* [Internet]. 2016 Dec 1 [cited 2018 Sep 18];17(1):144. Available from:

- <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1010-4>
144. Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, et al. Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat Protoc* [Internet]. 2014 Nov [cited 2018 Sep 18];9(11):2586–606. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25299156>
 145. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* [Internet]. 2014 Aug 14 [cited 2018 Sep 18];512(7513):155–60. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25079324>
 146. Almendro V, Cheng Y-K, Randles A, Itzkovitz S, Marusyk A, Ametller E, et al. Inference of Tumor Evolution during Chemotherapy by Computational Modeling and In Situ Analysis of Genetic and Phenotypic Cellular Diversity. *Cell Rep* [Internet]. 2014 Feb 13 [cited 2018 Sep 26];6(3):514–27. Available from: <https://www.sciencedirect.com/science/article/pii/S2211124713007997>
 147. Mehlen P, Puisieux A. Metastasis: a question of life or death. *Nat Rev Cancer*. 2006 Jun;6(6):449–58.
 148. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*. 2010 Oct;467(7319):1109–13.
 149. Chaffer CL, San Juan BP, Lim E, Weinberg RA. EMT, cell plasticity and metastasis. *Cancer Metastasis Rev* [Internet]. 2016 Dec 22 [cited 2018 Aug 24];35(4):645–54. Available from: <http://link.springer.com/10.1007/s10555-016-9648-7>
 150. Turajlic S, Swanton C. Metastasis as an evolutionary process. *Science* [Internet]. 2016 Apr 8 [cited 2018 Sep 14];352(6282):169–75. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27124450>
 151. Valastyan S, Weinberg RA. Tumor Metastasis: Molecular Insights and Evolving Paradigms. *Cell* [Internet]. 2011 Oct 14 [cited 2018 Sep 14];147(2):275–92. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22000009>
 152. Sethi N, Kang Y. Unravelling the complexity of metastasis — molecular understanding and targeted therapies. *Nat Rev Cancer* [Internet]. 2011 Oct 1 [cited 2018 Sep 14];11(10):735–48. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21941285>
 153. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-

- like breast cancer metastasis and xenograft. *Nature*. 2010;464(7291):999–1005.
154. Naxerova K, Jain RK. Using tumour phylogenetics to identify the roots of metastasis in humans. *Nat Rev Clin Oncol* [Internet]. 2015 May 20 [cited 2018 Sep 14];12(5):258–72. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25601447>
 155. Krøigård AB, Larsen MJ, Brasch-Andersen C, Lænkholm A-V, Knoop AS, Jensen JD, et al. Genomic Analyses of Breast Cancer Progression Reveal Distinct Routes of Metastasis Emergence. *Sci Rep* [Internet]. 2017 [cited 2018 Sep 14];7:43813. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28276460>
 156. McPherson A, Roth A, Laks E, Masud T, Bashashati A, Zhang AW, et al. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat Genet* [Internet]. 2016 Jul 16 [cited 2018 Sep 18];48(7):758–67. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27182968>
 157. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell* [Internet]. 2017 Aug [cited 2018 Sep 2];32(2):169–184.e7. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1535610817302970>
 158. Padua D, Zhang XH-F, Wang Q, Nadal C, Gerald WL, Gomis RR, et al. TGF β primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell*. 2008 Apr;133(1):66–77.
 159. Bruna A, Greenwood W, Le Quesne J, Teschendorff A, Miranda-Saavedra D, Rueda OM, et al. TGF β induces the formation of tumour-initiating cells in claudinlow breast cancer. *Nat Commun*. 2012;3:1055.
 160. Visvader JE, Lindeman GJ. Cancer Stem Cells: Current Status and Evolving Complexities. *Cell Stem Cell* [Internet]. 2017 Oct 28;10(6):717–28. Available from: <http://dx.doi.org/10.1016/j.stem.2012.05.007>
 161. Charafe-Jauffret E, Ginestier C, Iovino F, Wicinski J, Cervera N, Finetti P, et al. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Breast Cancer Res*. 2008 May;1(5):555–67.
 162. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci* . 2003 Apr;100(7):3983–8.
 163. Battula VL, Shi Y, Evans KW, Wang R-Y, Spaeth EL, Jacamo RO, et al. Ganglioside GD2

- identifies breast cancer stem cells and promotes tumorigenesis. *J Clin Invest*. 2012 Jun;122(6):2066–78.
164. Ghebeh H, Sleiman GM, Manogaran PS, Al-Mazrou A, Barhoush E, Al-Mohanna FH, et al. Profiling of normal and malignant breast tissue show CD44^{high}/CD24^{low} phenotype as a predominant stem/progenitor marker when used in combination with Ep-CAM/CD49f markers. *BMC Cancer*. 2013 Jun;13:289.
165. Cho RW, Wang X, Diehn M, Shedden K, Chen GY, Sherlock G, et al. Isolation and molecular characterization of cancer stem cells in MMTV-Wnt-1 murine breast tumors. *Stem Cells*. 2008 Feb;26(2):364–71.
166. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, et al. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell*. 2007 Nov;1(5):555–67.
167. Hwang-Verslues WW, Kuo W-H, Chang P-H, Pan C-C, Wang H-H, Tsai S-T, et al. Multiple lineages of human breast cancer stem/progenitor cells identified by profiling with stem cell markers. *PLoS One*. 2009 Dec;4(12):e8377.
168. Moraes RC, Zhang X, Harrington N, Fung JY, Wu M-F, Hilsenbeck SG, et al. Constitutive activation of smoothened (SMO) in mammary glands of transgenic mice leads to increased proliferation, altered differentiation and ductal dysplasia. *Development*. 2007 Mar;134(6):1231–42.
169. Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, et al. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev*. 2003 May;17(10):1253–70.
170. Grimshaw MJ, Cooper L, Papazisis K, Coleman JA, Bohnenkamp HR, Chiapero-Stanke L, et al. Mammosphere culture of metastatic breast cancer cells enriches for tumorigenic breast cancer cells. *Breast Cancer Res [Internet]*. 2008 Jun;10(3):R52. Available from: <https://doi.org/10.1186/bcr2106>
171. Ali HR, Dawson S-J, Blows FM, Provenzano E, Pharoah PD, Caldas C. Cancer stem cell markers in breast cancer: pathological, clinical and prognostic significance. *Breast Cancer Res [Internet]*. 2011 Nov;13(6):R118. Available from: <https://doi.org/10.1186/bcr3061>
172. Herschkowitz JI, Zhao W, Zhang M, Usary J, Murrow G, Edwards D, et al. Comparative oncogenomics identifies breast tumors enriched in functional tumor-initiating cells. *Proc Natl Acad Sci U S A*. 2012 Feb;109(8):2778–83.

173. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* 2010;12(5):R68.
174. Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc Natl Acad Sci U S A.* 2009 Aug;106(33):13820–5.
175. Fillmore CM, Kuperwasser C. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast Cancer Res.* 2008;10(2):R25.
176. Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu M-F, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. *J Natl Cancer Inst.* 2008 May;100(9):672–9.
177. Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C, et al. let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell.* 2007 Dec;131(6):1109–23.
178. Goodell MA, Brose K, Paradis G, Conner AS, Mulligan RC. Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *J Exp Med.* 1996 Apr;183(4):1797–806.
179. Hirschmann-Jax C, Foster AE, Wulf GG, Nuchtern JG, Jax TW, Gobel U, et al. A distinct “side population” of cells with high drug efflux capacity in human tumor cells. *Proc Natl Acad Sci U S A.* 2004 Sep;101(39):14228–33.
180. Patrawala L, Calhoun T, Schneider-Broussard R, Zhou J, Claypool K, Tang DG. Side population is enriched in tumorigenic, stem-like cancer cells, whereas ABCG2+ and ABCG2- cancer cells are similarly tumorigenic. *Cancer Res.* 2005 Jul;65(14):6207–19.
181. Zhou J, Wulfeuhle J, Zhang H, Gu P, Yang Y, Deng J, et al. Activation of the PTEN/mTOR/STAT3 pathway in breast cancer stem-like cells is required for viability and maintenance. *Proc Natl Acad Sci U S A.* 2007 Oct;104(41):16158–63.
182. Pece S, Tosoni D, Confalonieri S, Mazzarol G, Vecchi M, Ronzoni S, et al. Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. *Cell.* 2010 Jan;140(1):62–73.
183. Doyle LA, Yang W, Abruzzo L V, Krogmann T, Gao Y, Rishi AK, et al. A multidrug resistance transporter from human MCF-7 breast cancer cells. *Proc Natl Acad Sci U S A.* 1998 Dec;95(26):15665–70.

184. Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*. 2011 Aug;146(4):633–44.
185. Iliopoulos D, Hirsch HA, Wang G, Struhl K. Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. *Proc Natl Acad Sci U S A*. 2011 Jan;108(4):1397–402.
186. Meyer MJ, Fleming JM, Ali MA, Pesesky MW, Ginsburg E, Vonderhaar BK. Dynamic regulation of CD24 and the invasive, CD44posCD24neg phenotype in breast cancer cell lines. *Breast Cancer Res*. 2009;11(6):R82.
187. Mani SA, Guo W, Liao M-J, Eaton EN, Ayyanan A, Zhou AY, et al. The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. *Cell* [Internet]. 2008 May 16;133(4):704–15. Available from: <https://www.sciencedirect.com/science/article/pii/S0092867408004443>
188. Chaffer CL, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, et al. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proc Natl Acad Sci U S A*. 2011 May;108(19):7950–5.
189. Buczacki SJA, Zecchini HI, Nicholson AM, Russell R, Vermeulen L, Kemp R, et al. Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature*. 2013 Mar;495(7439):65–9.
190. Huch M, Dorrell C, Boj SF, Es JH Van, Wetering M Van De, Li VSW, et al. In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. *Nature*. 2013;494(7436):247–50.
191. Cheung TH, Rando T a. Molecular regulation of stem cell quiescence. *Nat Rev Mol Cell Biol*. 2013;14(6):329–40.
192. Kreso A, van Galen P, Pedley NM, Lima-Fernandes E, Frelin C, Davis T, et al. Self-renewal as a therapeutic target in human colorectal cancer. *Nat Med*. 2014 Jan;20(1):29–36.
193. Tomasetti C, Vogelstein B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Sci*. 2015 Jan;347(6217):78–81.
194. Easwaran H, Tsai H-CC, Baylin SBB. Cancer Epigenetics: Tumor Heterogeneity, Plasticity of Stem-like States, and Drug Resistance. *Mol Cell*. 2014 Jun;54(5):716–27.
195. Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature*. 2015 Nov;

196. Kreso A, O'Brien C a, van Galen P, Gan OI, Notta F, Brown AMK, et al. Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science*. 2013 Feb;339(6119):543–8.
197. Nguyen L V, Cox CL, Eirew P, Knapp DJHF, Pellacani D, Kannan N, et al. DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts. *Nat Commun*. 2014 Jan;5:5871.
198. Bruna A, Rueda OM, Greenwood W, Batra AS, Callari M, Batra RN, et al. A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. *Cell*. 2016 Sep;167(1):260–274.e22.
199. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. *Nat Methods* [Internet]. 2014 Apr 16 [cited 2018 Aug 22];11(4):396–8. Available from: <http://www.nature.com/articles/nmeth.2883>
200. Davis FM, Lloyd-Lewis B, Harris OB, Kozar S, Winton DJ, Muresan L, et al. Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. 2016 Oct 25;7:13053. Available from: <http://dx.doi.org/10.1038/ncomms13053>
201. Wu C, Li B, Lu R, Koelle SJ, Yang Y, Jares A, et al. Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. *Cell Stem Cell* [Internet]. 2014 Apr 3 [cited 2018 Aug 20];14(4):486–99. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24702997>
202. Lu R, Neff NF, Quake SR, Weissman IL. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat Biotechnol* [Internet]. 2011 Oct 2 [cited 2018 Aug 20];29(10):928–33. Available from: <http://www.nature.com/articles/nbt.1977>
203. Naik SH, Perié L, Swart E, Gerlach C, van Rooij N, de Boer RJ, et al. Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* [Internet]. 2013 Apr 3 [cited 2018 Aug 20];496(7444):229–32. Available from: <http://www.nature.com/articles/nature12013>
204. Porter SN, Baker LC, Mittelman D, Porteus MH. Lentiviral and targeted cellular barcoding reveals ongoing clonal dynamics of cell lines in vitro and in vivo. *Genome Biol* [Internet]. 2014 May 30 [cited 2018 Aug 25];15(5):R75. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24886633>

205. Hata AN, Niederst MJ, Archibald HL, Gomez-Caraballo M, Siddiqui FM, Mulvey HE, et al. Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat Med*. 2016 Mar;22(3):262–9.
206. Jakobsen KR, Demuth C, Madsen AT, Hussmann D, Vad-Nielsen J, Nielsen AL, et al. MET amplification and epithelial-to-mesenchymal transition exist as parallel resistance mechanisms in erlotinib-resistant, EGFR-mutated, NSCLC HCC827 cells. *Oncogenesis* [Internet]. 2017 Apr 3 [cited 2018 Sep 18];6(4):e307. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28368392>
207. Lopez Sambrooks C, Baro M, Quijano A, Narayan A, Cui W, Greninger P, et al. Oligosaccharyltransferase Inhibition Overcomes Therapeutic Resistance to EGFR Tyrosine Kinase Inhibitors. *Cancer Res* [Internet]. 2018 Sep 1 [cited 2018 Sep 18];78(17):5094–106. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30026325>
208. Cheong HT, Xu F, Choy CT, Hui CWC, Mok TSK, Wong CH. Upregulation of Bcl2 in NSCLC with acquired resistance to EGFR-TKI. *Oncol Lett* [Internet]. 2017 Nov 9 [cited 2018 Sep 18];15(1):901–7. Available from: <http://www.spandidos-publications.com/10.3892/ol.2017.7377>
209. Byers LA, Diao L, Wang J, Saintigny P, Girard L, Peyton M, et al. An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. *Clin Cancer Res*. 2013;19(1):279–90.
210. Yuan H, Wang Z, Gao C, Chen W, Huang Q, Yee J-K, et al. BCR-ABL Gene Expression Is Required for Its Mutations in a Novel KCL-22 Cell Culture Model for Acquired Resistance of Chronic Myelogenous Leukemia. *J Biol Chem* [Internet]. 2010 Feb 12 [cited 2018 Aug 25];285(7):5085–96. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20007699>
211. Huang L, Fu L. Mechanisms of resistance to EGFR tyrosine kinase inhibitors. *Acta Pharm Sin B* [Internet]. 2015 Sep 1 [cited 2018 Aug 25];5(5):390–401. Available from: <https://www.sciencedirect.com/science/article/pii/S2211383515001045>
212. Hata AN, Niederst MJ, Archibald HL, Gomez-Caraballo M, Siddiqui FM, Mulvey HE, et al. Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat Med* [Internet]. 2016 Mar 1 [cited 2018 Aug 25];22(3):262–9. Available from: <http://www.nature.com/articles/nm.4040>

213. Chang M. Tamoxifen Resistance in Breast Cancer. *Biomol Ther (Seoul)* [Internet]. 2012 May 6;20(3):256–67. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3794521/>
214. Knight WA, Livingston RB, Gregory EJ, McGuire WL. Estrogen Receptor as an Independent Prognostic Factor for Early Recurrence in Breast Cancer. *Cancer Res* [Internet]. 1977 Dec 1;37(12):4669 LP-4671. Available from: <http://cancerres.aacrjournals.org/content/37/12/4669.abstract>
215. Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schutz G, Umesono K, et al. The nuclear receptor superfamily: the second decade. *Cell*. 1995 Dec;83(6):835–9.
216. Sever R, Glass CK. Signaling by Nuclear Receptors. *Cold Spring Harb Perspect Biol* [Internet]. 2013 Mar;5(3):a016709. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3578364/>
217. Wu Q, Chambliss K, Umetani M, Mineo C, Shaul PW. Non-nuclear estrogen receptor signaling in the endothelium. *J Biol Chem*. 2011 Apr;286(17):14737–43.
218. Klinge CM. Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res* [Internet]. 2001 Jul 15;29(14):2905–19. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC55815/>
219. Nilsson S, Makela S, Treuter E, Tujague M, Thomsen J, Andersson G, et al. Mechanisms of estrogen action. *Physiol Rev*. 2001 Oct;81(4):1535–65.
220. Misiti S, Schomburg L, Yen PM, Chin WW. Expression and hormonal regulation of coactivator and corepressor genes. *Endocrinology*. 1998 May;139(5):2493–500.
221. Jordan VC, O'Malley BW. Selective Estrogen-Receptor Modulators and Antihormonal Resistance in Breast Cancer. *J Clin Oncol* [Internet]. 2007 Dec 20;25(36):5815–24. Available from: <https://doi.org/10.1200/JCO.2007.11.3886>
222. Feng Q, O'Malley BW. Nuclear Receptor Modulation - Role of Coregulators in Selective Estrogen Receptor Modulator (SERM) Actions. *Steroids* [Internet]. 2014 Nov 16;90:39–43. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4192004/>
223. Chen D, Ma H, Hong H, Koh SS, Huang SM, Schurter BT, et al. Regulation of transcription by a protein methyltransferase. *Science*. 1999 Jun;284(5423):2174–7.
224. Nagy L, Kao HY, Chakravarti D, Lin RJ, Hassig CA, Ayer DE, et al. Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase. *Cell*. 1997 May;89(3):373–80.

225. Spencer TE, Jenster G, Burcin MM, Allis CD, Zhou J, Mizzen CA, et al. Steroid receptor coactivator-1 is a histone acetyltransferase. *Nature*. 1997 Sep;389(6647):194–8.
226. Zwart W, Theodorou V, Kok M, Canisius S, Linn S, Carroll JS. Oestrogen receptor-co-factor-chromatin specificity in the transcriptional regulation of breast cancer. *EMBO J*. 2011 Oct;30(23):4764–76.
227. Lydon JP, O'Malley BW. Minireview: Steroid Receptor Coactivator-3: A Multifarious Coregulator in Mammary Gland Metastasis. *Endocrinology* [Internet]. 2011 Jan [cited 2018 Aug 23];152(1):19–25. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21047941>
228. Horlein AJ, Naar AM, Heinzel T, Torchia J, Gloss B, Kurokawa R, et al. Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature*. 1995 Oct;377(6548):397–404.
229. Chen JD, Evans RM. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature*. 1995 Oct;377(6548):454–7.
230. Wen YD, Perissi V, Staszewski LM, Yang WM, Krones A, Glass CK, et al. The histone deacetylase-3 complex contains nuclear receptor corepressors. *Proc Natl Acad Sci U S A*. 2000 Jun;97(13):7202–7.
231. Lloyd-Lewis B, Harris OB, Watson CJ, Davis FM. Mammary Stem Cells: Premise, Properties, and Perspectives. *Trends Cell Biol*. 2017 Aug;27(8):556–67.
232. Rios AC, Fu NY, Lindeman GJ, Visvader JE. In situ identification of bipotent stem cells in the mammary gland. *Nature*. 2014 Feb;506(7488):322–7.
233. Scheele CLGJ, Hannezo E, Muraro MJ, Zomer A, Langedijk NSM, van Oudenaarden A, et al. Identity and dynamics of mammary stem cells during branching morphogenesis. *Nature* [Internet]. 2017 Feb 16;542(7641):313–7. Available from: <http://dx.doi.org/10.1038/nature21046>
234. Asselin-Labat M-L, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC, et al. Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol*. 2007 Feb;9(2):201–9.
235. Manavathi B, Samanthapudi VSK, Gajulapalli VNR. Estrogen receptor coregulators and pioneer factors: the orchestrators of mammary gland cell fate and development [Internet]. Vol. 2, *Frontiers in Cell and Developmental Biology* . 2014. p. 34. Available from: <https://www.frontiersin.org/article/10.3389/fcell.2014.00034>

236. Jozwik KM, Carroll JS. Pioneer factors in hormone-dependent cancers. *Nat Rev Cancer*. 2012 May;12(6):381–5.
237. Kouros-Mehr H, Slorach EM, Sternlicht MD, Werb Z. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell*. 2006 Dec;127(5):1041–55.
238. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a critical determinant of Estrogen Receptor function and endocrine response. *Nat Genet* [Internet]. 2011 Jan 12;43(1):27–33. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3024537/>
239. Bernardo GM, Lozada KL, Miedler JD, Harburg G, Hewitt SC, Mosley JD, et al. FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. *Development*. 2010 Jun;137(12):2045–54.
240. Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet (London, England)*. 2011 Aug;378(9793):771–84.
241. Cronin-Fenton DP, Damkier P, Lash TL. Metabolism and transport of tamoxifen in relation to its effectiveness: new perspectives on an ongoing controversy. *Future Oncol* [Internet]. 2014 Jan;10(1):107–22. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4319217/>
242. Jeselsohn R, Cornwell M, Pun M, Buchwalter G, Nguyen M, Bango C, et al. Embryonic transcription factor SOX9 drives breast cancer endocrine resistance. *Proc Natl Acad Sci* [Internet]. 2017 May 30;114(22):E4482–91. Available from: <http://www.pnas.org/content/114/22/E4482.abstract>
243. Croxtall JD, McKeage K. Fulvestrant: a review of its use in the management of hormone receptor-positive metastatic breast cancer in postmenopausal women. *Drugs*. 2011 Feb;71(3):363–80.
244. Lee CI, Goodwin A, Wilcken N. Fulvestrant for hormone-sensitive metastatic breast cancer. *Cochrane database Syst Rev*. 2017 Jan;1:CD011093.
245. Fabian CJ. The what, why and how of aromatase inhibitors: hormonal agents for treatment and prevention of breast cancer. *Int J Clin Pract* [Internet]. 2007 Dec;61(12):2051–63. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2228389/>
246. Cuzick J, Sestak I, Baum M, Buzdar A, Howell A, Dowsett M, et al. Effect of anastrozole and

- tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. *Lancet Oncol*. 2010 Dec;11(12):1135–41.
247. Ciruelos E, Pascual T, Arroyo Vozmediano ML, Blanco M, Manso L, Parrilla L, et al. The therapeutic role of fulvestrant in the management of patients with hormone receptor-positive breast cancer. *The Breast* [Internet]. 2014 Jun 1 [cited 2018 Aug 22];23(3):201–8. Available from: <https://www.sciencedirect.com/science/article/pii/S0960977614000174>
 248. Ross-innes CS, Stark R, Teschendorff AE, Holmes K a, Raza H, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012;481(7381):389–93.
 249. Rastelli F, Crispino S. Factors predictive of response to hormone therapy in breast cancer. *Tumori* [Internet]. [cited 2018 Aug 23];94(3):370–83. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18705406>
 250. Liu H, Lee E-S, De Los Reyes A, Zapf JW, Craig Jordan V. Silencing and Reactivation of the Selective Estrogen Receptor Modulator-Estrogen Receptor Complex 1 [Internet]. Vol. 61, *CANCER RESEARCH*. 2001 [cited 2018 Aug 23]. Available from: <http://cancerres.aacrjournals.org/content/canres/61/9/3632.full.pdf>
 251. Saxena NK, Sharma D. Epigenetic Reactivation of Estrogen Receptor: Promising Tools for Restoring Response to Endocrine Therapy. *Mol Cell Pharmacol* [Internet]. 2010 [cited 2018 Aug 23];2(5):191–202. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21499573>
 252. Zhou Q, Atadja P, Davidson NE. Histone deacetylase inhibitor LBH589 reactivates silenced estrogen receptor alpha (ER) gene expression without loss of DNA hypermethylation. *Cancer Biol Ther* [Internet]. 2007 Jan [cited 2018 Aug 23];6(1):64–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17172825>
 253. Yang X, Phillips DL, Ferguson AT, Nelson WG, Herman JG, Davidson NE. Synergistic activation of functional estrogen receptor (ER)-alpha by DNA methyltransferase and histone deacetylase inhibition in human ER-alpha-negative breast cancer cells. *Cancer Res* [Internet]. 2001 Oct 1 [cited 2018 Aug 23];61(19):7025–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11585728>
 254. Munster PN, Thurn KT, Thomas S, Raha P, Lacevic M, Miller A, et al. A phase II study of the histone deacetylase inhibitor vorinostat combined with tamoxifen for the treatment of

- patients with hormone therapy-resistant breast cancer. *Br J Cancer* [Internet]. 2011 Jun 7 [cited 2018 Aug 23];104(12):1828–35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21559012>
255. Riggins RB, Schrecengost RS, Guerrero MS, Bouton AH. Pathways to Tamoxifen Resistance. 2008;256(1):1–24.
256. Schiff R, Massarweh SA, Shou J, Bharwani L, Mohsin SK, Osborne CK. Cross-talk between estrogen receptor and growth factor pathways as a molecular target for overcoming endocrine resistance. *Clin Cancer Res* [Internet]. 2004 Jan 1 [cited 2018 Aug 23];10(1 Pt 2):331S–6S. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14734488>
257. Chen M, Cui Y-K, Huang W-H, Man K, Zhang G-J. Phosphorylation of estrogen receptor α at serine 118 is correlated with breast cancer resistance to tamoxifen. *Oncol Lett* [Internet]. 2013 Jul [cited 2018 Aug 23];6(1):118–24. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23946788>
258. Massarweh S, Osborne CK, Creighton CJ, Qin L, Tsimelzon A, Huang S, et al. Tamoxifen resistance in breast tumors is driven by growth factor receptor signaling with repression of classic estrogen receptor genomic function. *Cancer Res* [Internet]. 2008 Feb 1 [cited 2018 Aug 23];68(3):826–33. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18245484>
259. Heldring N, Nilsson M, Buehrer B, Treuter E, Gustafsson J-A. Identification of tamoxifen-induced coregulator interaction surfaces within the ligand-binding domain of estrogen receptors. *Mol Cell Biol* [Internet]. 2004 Apr [cited 2018 Aug 23];24(8):3445–59. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15060164>
260. Joel PB, Smith J, Sturgill TW, Fisher TL, Blenis J, Lannigan DA. pp90rsk1 regulates estrogen receptor-mediated transcription through phosphorylation of Ser-167. *Mol Cell Biol* [Internet]. 1998 Apr [cited 2018 Aug 23];18(4):1978–84. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9528769>
261. Bunone G, Briand PA, Miksicek RJ, Picard D. Activation of the unliganded estrogen receptor by EGF involves the MAP kinase pathway and direct phosphorylation. *EMBO J* [Internet]. 1996 May 1 [cited 2018 Aug 23];15(9):2174–83. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8641283>
262. Shou J, Massarweh S, Osborne CK, Wakeling AE, Ali S, Weiss H, et al. Mechanisms of Tamoxifen Resistance: Increased Estrogen Receptor-HER2/neu Cross-Talk in ER/HER2-

- Positive Breast Cancer. JNCI J Natl Cancer Inst [Internet]. 2004 Jun 16 [cited 2018 Aug 23];96(12):926–35. Available from: <https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djh166>
263. Johnston S, Pippin J, Pivot X, Lichinitser M, Sadeghi S, Dieras V, et al. Lapatinib Combined With Letrozole Versus Letrozole and Placebo As First-Line Therapy for Postmenopausal Hormone Receptor–Positive Metastatic Breast Cancer. J Clin Oncol [Internet]. 2009 Nov 20 [cited 2018 Aug 23];27(33):5538–46. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19786658>
264. Martin MB, Franke TF, Stoica GE, Chambon P, Katzenellenbogen BS, Stoica BA, et al. A Role for Akt in Mediating the Estrogenic Functions of Epidermal Growth Factor and Insulin-Like Growth Factor I¹. Endocrinology [Internet]. 2000 Dec [cited 2018 Aug 23];141(12):4503–11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11108261>
265. Creighton CJ, Fu X, Hennessy BT, Casa AJ, Zhang Y, Gonzalez-Angulo AM, et al. Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER+ breast cancer. Breast Cancer Res [Internet]. 2010 Jun 22 [cited 2018 Aug 23];12(3):R40. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20569503>
266. Miller TW, Balko JM, Arteaga CL. Phosphatidylinositol 3-Kinase and Antiestrogen Resistance in Breast Cancer. J Clin Oncol [Internet]. 2011 Nov 20 [cited 2018 Aug 23];29(33):4452–61. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22010023>
267. Sommer S, Fuqua SA. Estrogen receptor and breast cancer. Semin Cancer Biol [Internet]. 2001 Oct [cited 2018 Aug 23];11(5):339–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11562176>
268. Murphy LC, Simon SL, Parkes A, Leygue E, Dotzlaw H, Snell L, et al. Altered expression of estrogen receptor coregulators during human breast tumorigenesis. Cancer Res [Internet]. 2000 Nov 15 [cited 2018 Aug 23];60(22):6266–71. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11103781>
269. Harigopal M, Heymann J, Ghosh S, Anagnostou V, Camp RL, Rimm DL. Estrogen receptor co-activator (AIB1) protein expression by automated quantitative analysis (AQUA) in a breast cancer tissue microarray and association with patient outcome. Breast Cancer Res Treat [Internet]. 2009 May 3 [cited 2018 Aug 23];115(1):77–85. Available from:

- <http://link.springer.com/10.1007/s10549-008-0063-9>
270. Redmond AM, Bane FT, Stafford AT, McIlroy M, Dillon MF, Crotty TB, et al. Coassociation of Estrogen Receptor and p160 Proteins Predicts Resistance to Endocrine Treatment; SRC-1 is an Independent Predictor of Breast Cancer Recurrence. *Clin Cancer Res* [Internet]. 2009 Mar 10 [cited 2018 Aug 23];15(6):2098–106. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19276281>
 271. Haugan Moi LL, Flågeng MH, Gandini S, Guerrieri-Gonzaga A, Bonanni B, Lazzeroni M, et al. Cancer Therapy: Clinical Effect of Low-Dose Tamoxifen on Steroid Receptor Coactivator 3/Amplified in Breast Cancer 1 in Normal and Malignant Human Breast Tissue. 2010 [cited 2018 Aug 23]; Available from: www.aacrjournals.org
 272. Hurtado A, Holmes K a, Geistlinger TR, Hutcheson IR, Nicholson RI, Brown M, et al. Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen. *Nature*. 2008;456(7222):663–6.
 273. Mohammed H, Russell IA, Stark R, Rueda OM, Hickey TE, Tarulli G a., et al. Progesterone receptor modulates ER α action in breast cancer. *Nature*. 2015;
 274. Wicha MS, Liu S, Dontu G. Cancer stem cells: An old idea - A paradigm shift. *Cancer Res*. 2006;66(4):1883–90.
 275. Clarke MF, Hass AT. Cancer Stem Cells. In: *Encyclopedia of Molecular Cell Biology and Molecular Medicine* [Internet]. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2006 [cited 2018 Aug 24]. Available from: <http://doi.wiley.com/10.1002/3527600906.mcb.200300130>
 276. Simões BM, Piva M, Iriondo O, Comaills V, López-Ruiz JA, Zabalza I, et al. Effects of estrogen on the proportion of stem cells in the breast. *Breast Cancer Res Treat* [Internet]. 2011 Aug 22 [cited 2018 Aug 24];129(1):23–35. Available from: <http://link.springer.com/10.1007/s10549-010-1169-4>
 277. Piva M, Domenici G, Iriondo O, Rábano M, Simões BM, Comaills V, et al. Sox2 promotes tamoxifen resistance in breast cancer cells. *EMBO Mol Med* [Internet]. 2014 Jan 1 [cited 2018 Aug 24];6(1):66–79. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24178749>
 278. Santini R, Pietrobono S, Pandolfi S, Montagnani V, D'Amico M, Penachioni JY, et al. SOX2 regulates self-renewal and tumorigenicity of human melanoma-initiating cells. *Oncogene*. 2014;(August 2013):1–12.

279. Schaefer SM, Segalada C, Cheng PF, Bonalli M, Parfejevs V, Levesque MP, et al. Sox2 is dispensable for primary melanoma and metastasis formation. *Oncogene* [Internet]. 2017 Aug 3 [cited 2018 Aug 24];36(31):4516–24. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28368416>
280. Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*. 2006;126(4):663–76.
281. Gwak JM, Kim M, Kim HJ, Jang MH, Park SY. Expression of embryonal stem cell transcription factors in breast cancer: Oct4 as an indicator for poor clinical outcome and tamoxifen resistance. *Oncotarget* [Internet]. 2017 May 30 [cited 2018 Aug 24];8(22):36305–18. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28422735>
282. Bhatt S, Stender JD, Joshi S, Wu G, Katzenellenbogen BS. OCT-4: a novel estrogen receptor- α collaborator that promotes tamoxifen resistance in breast cancer cells. *Oncogene* [Internet]. 2016 Nov 11 [cited 2018 Aug 24];35(44):5722–34. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27065334>
283. Gurel B, Ali TZ, Montgomery EA, Begum S, Hicks J, Goggins M, et al. NKX3.1 as a Marker of Prostatic Origin in Metastatic Tumors. *Am J Surg Pathol* [Internet]. 2010 Aug [cited 2018 Aug 24];34(8):1097–105. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20588175>
284. Jia Y, Zhou J, Luo X, Chen M, Chen Y, Wang J, et al. KLF4 overcomes tamoxifen resistance by suppressing MAPK signaling pathway and predicts good prognosis in breast cancer. *Cell Signal* [Internet]. 2018 Jan 1 [cited 2018 Aug 24];42:165–75. Available from: <https://www.sciencedirect.com/science/article/pii/S0898656817302590?via%3Dihub>
285. Stone A, Zotenko E, Locke WJ, Korbie D, Millar EKA, Pidsley R, et al. DNA methylation of oestrogen-regulated enhancers defines endocrine sensitivity in breast cancer. *Nat Commun* [Internet]. 2015 Dec 14 [cited 2018 Aug 25];6(1):7758. Available from: <http://www.nature.com/articles/ncomms8758>
286. Kashyap V, Rezende NC, Scotland KB, Shaffer SM, Persson JL, Gudas LJ, et al. Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the NANOG, OCT4, and SOX2 pluripotency transcription factors with polycomb repressive complexes and stem cell microRNAs. *Stem Cells Dev* [Internet]. 2009 Sep [cited 2018 Aug 25];18(7):1093–108. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19480567>
287. Schmidt R, Plath K. The roles of the reprogramming factors Oct4, Sox2 and Klf4 in

- resetting the somatic cell epigenome during induced pluripotent stem cell generation. *Genome Biol* [Internet]. 2012 Oct 22 [cited 2018 Aug 24];13(10):251. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23088445>
288. Wu Y, Zhang Z, Cenciarini ME, Proietti CJ, Amasino M, Hong T, et al. Tamoxifen Resistance in Breast Cancer Is Regulated by the EZH2-ER α -GREB1 Transcriptional Axis. *Cancer Res* [Internet]. 2018 Feb 1 [cited 2018 Aug 25];78(3):671–84. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29212856>
 289. Cassidy JW, Batra AS, Greenwood W, Bruna A. Patient-derived tumour xenografts for breast cancer drug discovery. *Endocr Relat Cancer*. 2016 Dec;23(12):T259–70.
 290. Soule HD, Vazquez J, Long A, Albert S, Brennan M. A Human Cell Line From a Pleural Effusion Derived From a Breast Carcinoma 1,2. 1973;51(5):1409–16.
 291. Lacroix M, Leclercq G. Relevance of Breast Cancer Cell Lines as Models for Breast Tumours: An Update. *Breast Cancer Res Treat*. 2004;83(3):249–89.
 292. Carter P, Presta L, Gorman CM, Ridgway JB, Henner D, Wong WL, et al. Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc Natl Acad Sci U S A*. 1992 May;89(10):4285–9.
 293. Heiser LM, Sadanandam A, Kuo W-L, Benz SC, Goldstein TC, Ng S, et al. Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc Natl Acad Sci U S A*. 2012 Feb;109(8):2724–9.
 294. Polyak K. Molecular alterations in ductal carcinoma in situ of the breast. *Curr Opin Oncol*. 2002;14(1):92–6.
 295. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature*. 2001 Nov;414(6859):105–11.
 296. Tufegdžic-Vidakovic A, Rueda OM, Vervoort SJ, Batra AS, Goldgraben MA, Uribe-Lewis S, et al. Context-Specific Effects of TGF- β / SMAD3 in Cancer Are Modulated by the Epigenome Article Context-Specific Effects of TGF- β / SMAD3 in Cancer Are Modulated by the Epigenome. *CellReports*. 2015;13(11):2480–90.
 297. Cardi RD, Anver MR, Gusterson BA, Hennighausen L, Jensen RA, Merino MJ, et al. The mammary pathology of genetically engineered mice: the consensus report and recommendations from the Annapolis meeting { . 2000;968–88.
 298. Vargo-gogola T, Rosen JM. Modelling breast cancer: one size does not fit all. 2007;7(september):659–72.

299. Wyckoff JB, Wang Y, Lin EY, Li J, Goswami S, Stanley ER, et al. Direct Visualization of Macrophage-Assisted Tumor Cell Intravasation in Mammary Tumors. 2007;(6):2649–57.
300. Medina D, Thompson HJ. A Comparison of the Salient Features of Mouse, Rat, and Human Mammary Tumorigenesis. In: Ip MM, Asch BB, editors. *Methods in Mammary Gland Biology and Breast Cancer Research*. Boston, MA: Springer US; 2000. p. 31–6.
301. Dabydeen SA, Furth PA. Genetically engineered ER α -positive breast cancer mouse models. *Endocrine-Related Cancer*. 2014 Jun;21(3):R195–208.
302. Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature*. 2009 May;459(7244):262–5.
303. Huch M, Bonfanti P, Boj SF, Sato T, Loomans CJM, Wetering M Van De, et al. Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5 / R-spondin axis. *EMBO J*. 2013;32(20):2708–21.
304. Huch M, Gehart H, van Boxtel R, Hamer K, Blokzijl F, Verstegen MM, et al. Long-Term Culture of Genome-Stable Bipotent Stem Cells from Adult Human Liver. *Cell*. 2015 Jan;160(1–2):299–312.
305. van de Wetering M, Francies HE, Francis JM, Bounova G, Iorio F, Pronk A, et al. Prospective Derivation of a Living Organoid Biobank of Colorectal Cancer Patients. *Cell*. 2015 May;161(4):933–45.
306. Bergamaschi A, Hjortland GO, Triulzi T, Sørli T, Johnsen H, Ree AH, et al. Molecular profiling and characterization of luminal-like and basal-like in vivo breast cancer xenograft models. *Mol Oncol*. 2009 Dec;3(5–6):469–82.
307. DeRose YS, Wang G, Lin Y-C, Bernard PS, Buys SS, Ebbert MTW, et al. Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nat Med*. 2011 Jan;17(11):1514–20.
308. Siolas D, Hannon GJ. Patient Derived Tumor Xenografts: transforming clinical samples into mouse models. *Cancer Res*. 2013 Sep;73(17):5315–9.
309. Zhang X, Claerhout S, Prat A, Dobrolecki LE, Petrovic I, Lai Q, et al. A Renewable Tissue Resource of Phenotypically Stable, Biologically and Ethnically Diverse, Patient-Derived Human Breast Cancer Xenograft Models. *Cancer Res*. 2013 Jul;73(15):4885–97.
310. Sflomos G, Dormoy V, Metsalu T, Jeitziner R, Battista L, Scabia V, et al. A Preclinical Model for ER α -Positive Breast Cancer Points to the Epithelial Microenvironment as

- Determinant of Luminal Phenotype and Hormone Response. *Cancer Cell*. 2016 Mar;29(3):407–22.
311. Shi H, Hugo W, Kong X, Hong A, Koya RC, Moriceau G, et al. Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. *Cancer Discov*. 2014;4(1):1–15.
 312. Friedman AA, Amzallag A, Pruteanu-malinici I, Baniya S, Frederick T, Lawrence DP, et al. Landscape of Targeted Anti-Cancer Drug Synergies in Melanoma Identifies a Novel BRAF-VEGFR / PDGFR Combination Treatment. 2015;1–21.
 313. HUXLEY J. CANCER BIOLOGY: COMPARATIVE and GENETIC. *Biol Rev* [Internet]. 1956 Nov 1 [cited 2018 Aug 25];31(4):474–513. Available from: <http://doi.wiley.com/10.1111/j.1469-185X.1956.tb01558.x>
 314. Muto T, Bussey HJ, Morson BC. The evolution of cancer of the colon and rectum. *Cancer* [Internet]. 1975 Dec [cited 2018 Aug 25];36(6):2251–70. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1203876>
 315. Mcgranahan N, Swanton C. Leading Edge Review Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. 2017 [cited 2018 Aug 25]; Available from: <http://dx.doi.org/10.1016/j.cell.2017.01.018>
 316. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA Cancer J Clin* [Internet]. 2017 Nov 1 [cited 2018 Aug 25];67(6):439–48. Available from: <http://doi.wiley.com/10.3322/caac.21412>
 317. Coser KR, Wittner BS, Rosenthal NF, Collins SC, Melas A, Smith SL, et al. Antiestrogen-resistant subclones of MCF-7 human breast cancer cells are derived from a common monoclonal drug-resistant progenitor. *Proc Natl Acad Sci* [Internet]. 2009 Oct 25 [cited 2018 Aug 22];106(34):14536–41. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2823256>
 318. World Health Organization. WHO List of Essential Medicines [Internet]. 2018 [cited 2018 Aug 25]. Available from: <http://www.who.int/medicines/publications/essentialmedicines/en/>
 319. Yuan J, Liu M, Yang L, Tu G, Zhu Q, Chen M, et al. Acquisition of epithelial-mesenchymal transition phenotype in the tamoxifen-resistant breast cancer cell: a new role for G protein-coupled estrogen receptor in mediating tamoxifen resistance through cancer-associated fibroblast-derived fibronectin and β 1-integrin signaling pathway in tumor cells. *Breast Cancer Res* [Internet]. 2015 May 21 [cited 2018 Aug 22];17(1):69. Available

- from: <http://www.ncbi.nlm.nih.gov/pubmed/25990368>
320. Riggins RB, Schrecengost RS, Guerrero MS, Bouton AH. Pathways to tamoxifen resistance. *Cancer Lett* [Internet]. 2007 Oct 18 [cited 2018 Aug 23];256(1):1–24. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17475399>
 321. Notas G, Pelekanou V, Kampa M, Alexakis K, Sfakianakis S, Laliotis A, et al. Tamoxifen induces a pluripotency signature in breast cancer cells and human tumors. *Mol Oncol* [Internet]. 2015;9(9):1744–59. Available from: <http://www.sciencedirect.com/science/article/pii/S1574789115001234>
 322. King M, Pearson T, Shultz LD, Leif J, Bottino R, Trucco M, et al. A new Hu-PBL model for the study of human islet alloreactivity based on NOD-scid mice bearing a targeted mutation in the IL-2 receptor gamma chain gene. *Clin Immunol* [Internet]. 2008 Mar [cited 2018 Sep 18];126(3):303–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18096436>
 323. Augsten M. Cancer-associated fibroblasts as another polarized cell type of the tumor microenvironment. *Front Oncol*. 2014 Jan;4(March):62.
 324. Hu Y, Smyth GK. ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J Immunol Methods* [Internet]. 2009 Aug 15 [cited 2018 Aug 30];347(1–2):70–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19567251>
 325. Knowlden JM, Hutcheson IR, Jones HE, Madden T, Gee JMW, Harper ME, et al. Elevated levels of epidermal growth factor receptor/c-erbB2 heterodimers mediate an autocrine growth regulatory pathway in tamoxifen-resistant MCF-7 cells. *Endocrinology*. 2003;144(3):1032–44.
 326. Maximov PY, McDaniel RE, Fernandes DJ, Korostyshevskiy VR, Bhatta P, Mürdter TE, et al. Simulation with cells in vitro of tamoxifen treatment in premenopausal breast cancer patients with different CYP2D6 genotypes. *Br J Pharmacol* [Internet]. 2014 Dec [cited 2018 Sep 1];171(24):5624–35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25073551>
 327. Debnath J, Muthuswamy SK, Brugge JS. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* [Internet]. 2003 Jul 1 [cited 2018 Aug 30];30(3):256–68. Available from: <https://www.sciencedirect.com/science/article/pii/S104620230300032X?via%3Dihub>

328. Welm BE, Dijkgraaf GJP, Bledau AS, Welm AL, Werb Z. Lentiviral Transduction of Mammary Stem Cells for Analysis of Gene Function during Development and Cancer. *Cell Stem Cell* [Internet]. 2008 Jan 10 [cited 2018 Aug 26];2(1):90–102. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18371425>
329. Papapetrou EP, Tomishima MJ, Chambers SM, Mica Y, Reed E, Menon J, et al. Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. *Proc Natl Acad Sci* [Internet]. 2009 Aug 4 [cited 2018 Aug 26];106(31):12759–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19549847>
330. Dull T, Zufferey R, Kelly M, Mandel RJ, Nguyen M, Trono D, et al. A third-generation lentivirus vector with a conditional packaging system. *J Virol* [Internet]. 1998 Nov [cited 2018 Sep 1];72(11):8463–71. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9765382>
331. Stewart SA, Dykxhoorn DM, Palliser D, Mizuno H, Yu EY, An DS, et al. Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* [Internet]. 2003 Apr [cited 2018 Sep 1];9(4):493–501. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12649500>
332. Ichim C V, Wells R a. Generation of high-titer viral preparations by concentration using successive rounds of ultracentrifugation. *J Transl Med*. 2011;9(1):137.
333. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc* [Internet]. 2008 Jun 1 [cited 2018 Sep 1];3(6):1101–8. Available from: <http://www.nature.com/articles/nprot.2008.73>
334. Bhang HC, Ruddy DA, Krishnamurthy Radhakrishna V, Caushi JX, Zhao R, Hims MM, et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med* [Internet]. 2015 Apr 13 [cited 2018 Aug 20];21(5):440–8. Available from: <http://www.nature.com/doifinder/10.1038/nm.3841>
335. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. *Nat Methods*. 2014;11(4):396–8.
336. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, et al. Comprehensive molecular portraits of human breast tumours. *Nature* [Internet]. 2012 Sep 23 [cited 2018 Aug 26];490(7418):61–70. Available from: <http://www.nature.com/doifinder/10.1038/nature11412>
337. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer

- Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401–4.
338. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci Signal.* 2013 Apr;6(269):pl1-pl1.
339. Huang M, Shen A, Ding J, Geng M. Molecularly targeted cancer therapy: some lessons from the past decade. *Trends Pharmacol Sci.* 2014 Jan;35(1):41–50.
340. Albino a P, Le Strange R, Oliff a I, Furth ME, Old LJ. Transforming ras genes from human melanoma: a manifestation of tumour heterogeneity? *Nature.* 1984;308(5954):69–72.
341. Dontu G, El-Ashry D, Wicha MS. Breast cancer, stem/progenitor cells and the estrogen receptor. *Trends Endocrinol Metab.* 2004;15(5):193–7.
342. Sim BM, Piva M, Domenici G, Iriondo O, Miriam R, Zabalza I, et al. Sox 2 promotes tamoxifen resistance in breast cancer cells. 2014;6(1):66–79.
343. Nawata H, Bronzert D, Lippman ME. Isolation and characterization of a tamoxifen-resistant cell line derived from MCF-7 human breast cancer cells. *J Biol Chem.* 1981;256(10):5016–21.
344. Gutierrez MC, Detre S, Johnston S, Mohsin SK, Shou J, Allred DC, et al. Molecular changes in tamoxifen-resistant breast cancer: Relationship between estrogen receptor, HER-2, and p38 mitogen-activated protein kinase. *J Clin Oncol.* 2005;23(11):2469–76.
345. Leung E, Kannan N, Krissansen GW, Findlay MP, Baguley BC. MCF-7 breast cancer cells selected for tamoxifen resistance acquire new phenotypes differing in DNA content, phospho-HER2 and PAX2 expression, and rapamycin sensitivity. *Cancer Biol Ther.* 2010;9(9):717–24.
346. Ling GQ, Chen DB, Wang BQ, Zhang LS. Expression of the pluripotency markers Oct3/4, Nanog and Sox2 in human breast cancer cell lines. *Oncol Lett.* 2012;4(6):1264–8.
347. Gerrits A, Dykstra B, Kalmykova OJ, Klauke K, Verovskaya E, Broekhuis MJC, et al. Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood.* 2010;115(13):2610–8.
348. Strober W. Trypan Blue Exclusion Test of Cell Viability. In: *Current Protocols in Immunology* [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2001 [cited 2018 Aug 20]. p. Appendix 3B. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18432654>
349. Sasson S. Equilibrium binding analysis of estrogen agonists and antagonists: relation to

- the activation of the estrogen receptor. *Pathol Biol (Paris)* [Internet]. 1991 Jan [cited 2018 Aug 20];39(1):59–69. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2011412>
350. Sasson S, Notides AC. Mechanism of the Estrogen Receptor Interaction with 4-Hydroxytamoxifen. *Mol Endocrinol* [Internet]. 1988 Apr [cited 2018 Aug 20];2(4):307–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/3380103>
 351. Wang D-Y, Fulthorpe R, Liss SN, Edwards EA. Identification of Estrogen-Responsive Genes by Complementary Deoxyribonucleic Acid Microarray and Characterization of a Novel Early Estrogen-Induced Gene: *EEIG1*. *Mol Endocrinol* [Internet]. 2004 Feb 1 [cited 2018 Aug 22];18(2):402–11. Available from: <https://academic.oup.com/mend/article-lookup/doi/10.1210/me.2003-0202>
 352. Tilli TM, Castro C da S, Tuszynski JA, Carels N. A strategy to identify housekeeping genes suitable for analysis in breast cancer diseases. *BMC Genomics* [Internet]. 2016 [cited 2018 Sep 1];17(1):639. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27526934>
 353. Kastner P, Krust A, Turcotte B, Stropp U, Tora L, Gronemeyer H, et al. Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. *EMBO J* [Internet]. 1990 May [cited 2018 Aug 20];9(5):1603–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2328727>
 354. Artymovich K, Appledorn DM. A Multiplexed Method for Kinetic Measurements of Apoptosis and Proliferation Using Live-Content Imaging. In: *Methods in molecular biology* (Clifton, NJ) [Internet]. 2015 [cited 2018 Sep 1]. p. 35–42. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25308260>
 355. Berthois Y, Katzenellenbogen JA, Katzenellenbogen BS. Phenol red in tissue culture media is a weak estrogen: implications concerning the study of estrogen-responsive cells in culture. *Proc Natl Acad Sci U S A* [Internet]. 1986 Apr [cited 2018 Aug 20];83(8):2496–500. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/3458212>
 356. Sikora MJ, Johnson MD, Lee A V, Oesterreich S. Endocrine Response Phenotypes Are Altered by Charcoal-Stripped Serum Variability. *Endocrinology* [Internet]. 2016 Oct [cited 2018 Aug 20];157(10):3760–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27459541>
 357. Weisberg E, Manley PW, Cowan-Jacob SW, Hochhaus A, Griffin JD. Second generation inhibitors of BCR-ABL for the treatment of imatinib-resistant chronic myeloid leukaemia.

- Nat Rev Cancer [Internet]. 2007 May 1 [cited 2018 Aug 20];7(5):345–56. Available from: <http://www.nature.com/articles/nrc2126>
358. Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. Nat Rev Cancer [Internet]. 2013 Oct 1 [cited 2018 Aug 20];13(10):714–26. Available from: <http://www.nature.com/articles/nrc3599>
 359. Ohashi K, Maruvka YE, Michor F, Pao W. Epidermal growth factor receptor tyrosine kinase inhibitor-resistant disease. J Clin Oncol [Internet]. 2013 Mar 10 [cited 2018 Aug 20];31(8):1070–80. Available from: <http://ascopubs.org/doi/10.1200/JCO.2012.43.3912>
 360. Figliozzi RW, Chen F, Chi A, Hsia S-CV. Using the inverse Poisson distribution to calculate multiplicity of infection and viral replication by a high-throughput fluorescent imaging system. Virol Sin [Internet]. 2016 Apr [cited 2018 Aug 22];31(2):180–3. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26826079>
 361. Geraerts M, Willems S, Baekelandt V, Debyser Z, Gijsbers R. Comparison of lentiviral vector titration methods. BMC Biotechnol [Internet]. 2006 Jul 12 [cited 2018 Aug 21];6(1):34. Available from: <http://bmcbiotechnol.biomedcentral.com/articles/10.1186/1472-6750-6-34>
 362. Ector Hern Andez-Vargas H, Ballestar E, Carmona-Saez P, Von Kobbe C, Bañón-Rodríguez I, Esteller M, et al. Transcriptional profiling of MCF7 breast cancer cells in response to 5-Fluorouracil: Relationship with cell cycle changes and apoptosis, and identification of novel targets of p53. Int J Cancer [Internet]. 2006 [cited 2018 Aug 21];119:1164–75. Available from: <http://www.interscience.wiley.com/jpages/0020-7136/suppmat>.
 363. Jost L. *Entropy and diversity*. Oikos [Internet]. 2006 May 1 [cited 2018 Sep 2];113(2):363–75. Available from: <http://doi.wiley.com/10.1111/j.2006.0030-1299.14714.x>
 364. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, et al. LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1. Science (80-) [Internet]. 2003 Oct 17 [cited 2018 Aug 21];302(5644):415–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14564000>
 365. Biffi A, Bartolomae CC, Cesana D, Cartier N, Aubourg P, Ranzani M, et al. Lentiviral vector common integration sites in preclinical models and a clinical trial reflect a benign integration bias and not oncogenic selection. Blood [Internet]. 2011 May 19 [cited 2018 Aug 21];117(20):5332–9. Available from: <http://www.bloodjournal.org/content/117/20/5332.full>

- <http://www.ncbi.nlm.nih.gov/pubmed/21403130>
366. Faronato M, Nguyen VTM, Patten DK, Lombardo Y, Steel JH, Patel N, et al. DMXL2 drives epithelial to mesenchymal transition in hormonal therapy resistant breast cancer through Notch hyper-activation. *Oncotarget* [Internet]. 2015 Sep 8 [cited 2018 Sep 24];6(26):22467–79. Available from: <http://www.oncotarget.com/fulltext/4164>
 367. Tonetti DA, Jordan VC. The role of estrogen receptor mutations in tamoxifen-stimulated breast cancer. *J Steroid Biochem Mol Biol* [Internet]. 1997 Jun [cited 2018 Aug 22];62(2–3):119–28. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9393947>
 368. Fischer A, Vázquez-García I, Illingworth CJR, Mustonen V. High-definition reconstruction of clonal composition in cancer. *Cell Rep*. 2014;7:1740–52.
 369. Shi Y, Yan H, Frost P, Gera J, Lichtenstein A. Mammalian target of rapamycin inhibitors activate the AKT kinase in multiple myeloma cells by up-regulating the insulin-like growth factor receptor/insulin receptor substrate-1/phosphatidylinositol 3-kinase cascade. *Mol Cancer Ther* [Internet]. 2005 Oct 1 [cited 2018 Aug 22];4(10):1533–40. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16227402>
 370. Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA, et al. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell*. 1998 Dec;95(7):927–37.
 371. Prinz H. Hill coefficients, dose-response curves and allosteric mechanisms. *J Chem Biol* [Internet]. 2010 Mar [cited 2018 Aug 22];3(1):37–44. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19779939>
 372. Maguire JJ, Kuc RE, Davenport AP. Radioligand Binding Assays and Their Analysis. In: *Methods in molecular biology* (Clifton, NJ) [Internet]. 2012 [cited 2018 Sep 2]. p. 31–77. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22674160>
 373. Genestie C, Zafrani B, Asselain B, Fourquet A, Rozan S, Validire P, et al. Comparison of the prognostic value of Scarff-Bloom-Richardson and Nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems. *Anticancer Res* [Internet]. [cited 2018 Sep 23];18(1B):571–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9568179>
 374. Ciriello G. Comprehensive molecular portraits of invasive lobular breast. *Cancer Cell*. 2015;163:506–19.
 375. Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis

- of the randomised trials. *Lancet* [Internet]. 2015 Oct 3 [cited 2018 Sep 23];386(10001):1341–52. Available from: <https://www.sciencedirect.com/science/article/pii/S0140673615610741?via%3Dihub>
376. Patten DK, Corleone G, Györfy B, Perone Y, Slaven N, Barozzi I, et al. Enhancer mapping uncovers phenotypic heterogeneity and evolution in patients with luminal breast cancer. *Nat Med* [Internet]. 2018 Sep 23 [cited 2018 Sep 21];24(9):1469–80. Available from: <http://www.nature.com/articles/s41591-018-0091-x>
377. Magnani L, Eeckhoutte J, Lupien M. Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends Genet* [Internet]. 2011 Nov 1 [cited 2018 Sep 23];27(11):465–74. Available from: <https://www.sciencedirect.com/science/article/pii/S0168952511001107?via%3Dihub>
378. Falahi F, Huisman C, Kazemier HG, van der Vlies P, Kok K, Hospers GAP, et al. Towards sustained silencing of HER2/neu in cancer by epigenetic editing. *Mol Cancer Res* [Internet]. 2013 Sep 1 [cited 2018 Sep 23];11(9):1029–39. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23814024>
379. Laprell F, Finkl K, Müller J. Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. *Science* [Internet]. 2017 Apr 7 [cited 2018 Sep 23];356(6333):85–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28302792>
380. Wang X, Moazed D. DNA sequence-dependent epigenetic inheritance of gene silencing and histone H3K9 methylation. *Science* [Internet]. 2017 Apr 7 [cited 2018 Sep 23];356(6333):88–91. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28302794>
381. Coleman RT, Struhl G. Causal role for inheritance of H3K27me3 in maintaining the OFF state of a *Drosophila* HOX gene. *Science* [Internet]. 2017 Apr 7 [cited 2018 Sep 23];356(6333):eaai8236. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28302795>
382. Magnani L, Frigè G, Gadaleta RM, Corleone G, Fabris S, Kempe H, et al. Acquired CYP19A1 amplification is an early specific mechanism of aromatase inhibitor resistance in ERα metastatic breast cancer. *Nat Genet* [Internet]. 2017 Mar 23 [cited 2018 Sep 23];49(3):444–50. Available from: <http://www.nature.com/articles/ng.3773>
383. Greaves M, Maley CC. Clonal evolution in cancer. *Nature* [Internet]. 2012 Jan 19 [cited 2018 Aug 22];481(7381):306–13. Available from:

- <http://www.ncbi.nlm.nih.gov/pubmed/22258609>
384. Foo J, Leder K, Mumenthaler SM. Cancer as a moving target: understanding the composition and rebound growth kinetics of recurrent tumors. *Evol Appl* [Internet]. 2013 Jan [cited 2018 Aug 22];6(1):54–69. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23396647>
 385. Robasky K, Lewis NE, Church GM. The role of replicates for error mitigation in next-generation sequencing. *Nat Rev Genet* [Internet]. 2014 Jan 10 [cited 2018 Aug 20];15(1):56–62. Available from: <http://www.nature.com/articles/nrg3655>
 386. Toy W, Shen Y, Won H, Green B, Sakr RA, Will M, et al. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet* [Internet]. 2013 Dec 3 [cited 2018 Sep 2];45(12):1439–45. Available from: <http://www.nature.com/articles/ng.2822>
 387. Fanning SW, Mayne CG, Dharmarajan V, Carlson KE, Martin TA, Novick SJ, et al. Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation. *Elife* [Internet]. 2016 Feb 2 [cited 2018 Sep 2];5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26836308>
 388. Jeselsohn R, Yelensky R, Buchwalter G, Frampton G, Meric-Bernstam F, Gonzalez-Angulo AM, et al. Emergence of Constitutively Active Estrogen Receptor- Mutations in Pretreated Advanced Estrogen Receptor-Positive Breast Cancer. *Clin Cancer Res* [Internet]. 2014 Apr 1 [cited 2018 Sep 2];20(7):1757–67. Available from: <http://clincancerres.aacrjournals.org/cgi/doi/10.1158/1078-0432.CCR-13-2332>
 389. Schiavon G, Hrebien S, Garcia-Murillas I, Cutts RJ, Pearson A, Tarazona N, et al. Analysis of *ESR1* mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Sci Transl Med* [Internet]. 2015 Nov 11 [cited 2018 Sep 2];7(313):313ra182–313ra182. Available from: <http://stm.sciencemag.org/lookup/doi/10.1126/scitranslmed.aac7551>
 390. Fribbens C, O’Leary B, Kilburn L, Hrebien S, Garcia-Murillas I, Beaney M, et al. Plasma *ESR1* Mutations and the Treatment of Estrogen Receptor–Positive Advanced Breast Cancer. *J Clin Oncol* [Internet]. 2016 Sep [cited 2018 Sep 2];34(25):2961–8. Available from: <http://ascopubs.org/doi/10.1200/JCO.2016.67.3061>
 391. Chandarlapaty S, Chen D, He W, Sung P, Samoila A, You D, et al. Prevalence of *ESR1* Mutations in Cell-Free DNA and Outcomes in Metastatic Breast Cancer. *JAMA Oncol*

- [Internet]. 2016 Oct 1 [cited 2018 Sep 2];2(10):1310. Available from: <http://oncology.jamanetwork.com/article.aspx?doi=10.1001/jamaoncol.2016.1279>
392. O'Leary B, Hrebien S, Morden JP, Beaney M, Fribbens C, Huang X, et al. Early circulating tumor DNA dynamics and clonal selection with palbociclib and fulvestrant for breast cancer. *Nat Commun* [Internet]. 2018 Dec 1 [cited 2018 Sep 2];9(1):896. Available from: <http://www.nature.com/articles/s41467-018-03215-x>
 393. Hiscox S, Jiang WG, Obermeier K, Taylor K, Morgan L, Burmi R, et al. Tamoxifen resistance in MCF7 cells promotes EMT-like behaviour and involves modulation of β -catenin phosphorylation. *Int J Cancer* [Internet]. 2006 Jan 15 [cited 2018 Aug 24];118(2):290–301. Available from: <http://doi.wiley.com/10.1002/ijc.21355>
 394. Wang D, Lu P, Zhang H, Luo M, Zhang X, Wei X, et al. Oct-4 and Nanog promote the epithelial-mesenchymal transition of breast cancer stem cells and are associated with poor prognosis in breast cancer patients. 2014;5(21).
 395. Dubrovskaya a, Hartung a, Bouchez LC, Walker JR, Reddy V a, Cho CY, et al. CXCR4 activation maintains a stem cell population in tamoxifen-resistant breast cancer cells through AhR signalling. *Br J Cancer*. 2012;107(1):43–52.
 396. Wang Q, Jiang J, Ying G, Xie X-Q, Zhang X, Xu W, et al. Tamoxifen enhances stemness and promotes metastasis of ER α 36+ breast cancer by upregulating ALDH1A1 in cancer cells. *Cell Res* [Internet]. 2018 Mar [cited 2018 Aug 26];28(3):336–58. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29393296>
 397. Fu H, Fu L, Xie C, Zuo W-S, Liu Y-S, Zheng M-Z, et al. miR-375 inhibits cancer stem cell phenotype and tamoxifen resistance by degrading HOXB3 in human ER-positive breast cancer. *Oncol Rep* [Internet]. 2017 Jan 1 [cited 2018 Aug 26];37(2):1093–9. Available from: <https://www.spandidos-publications.com/10.3892/or.2017.5360>
 398. Yamaguchi N, Nakayama Y, Yamaguchi N. Involvement of FOXA1 down-regulation in CSC-like properties Down-regulation of Forkhead Box Protein A1 (FOXA1) Leads to Cancer-stem Cell-like Properties in Tamoxifen-resistant Breast Cancer Cells through Induction of Interleukin-6 Running title: Involvement of FOXA1 down-regulation in CSC-like properties From. 2017 [cited 2018 Aug 26]; Available from: <http://www.jbc.org/cgi/doi/10.1074/jbc.M116.763276>
 399. Ponti D, Zaffaroni N, Capelli C, Daidone MG. Breast cancer stem cells: An overview. *Eur J Cancer*. 2006;42(9):1219–24.

400. Scheel C, Weinberg R a. Cancer stem cells and epithelial-mesenchymal transition: Concepts and molecular links. *Semin Cancer Biol.* 2012;22(5-6):396-403.
401. Nguyen VTM, Barozzi I, Faronato M, Lombardo Y, Steel JH, Patel N, et al. Differential epigenetic reprogramming in response to specific endocrine therapies promotes cholesterol biosynthesis and cellular invasion. *Nat Commun [Internet]*. 2015 Dec 27 [cited 2018 Sep 24];6(1):10044. Available from: <http://www.nature.com/articles/ncomms10044>
402. Sheridan C, Kishimoto H, Fuchs RK, Mehrotra S, Bhat-Nakshatri P, Turner CH, et al. CD44+/CD24- breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis. *Breast Cancer Res.* 2006;8(5):R59.
403. Phillips TM, McBride WH, Pajonk F. The response of CD24(-/low)/CD44+ breast cancer-initiating cells to radiation. *J Natl Cancer Inst.* 2006 Dec;98(24):1777-85.
404. Pastrana E, Silva-Vargas V, Doetsch F. Eyes Wide Open: A Critical Review of Sphere-Formation as an Assay for Stem Cells. *Cell Stem Cell [Internet]*. 2011 May 6 [cited 2018 Aug 30];8(5):486-98. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21549325>
405. Reynolds BA, Weiss S. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science [Internet]*. 1992 Mar 27 [cited 2018 Aug 30];255(5052):1707-10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1553558>
406. Ponti D, Costa A, Zaffaroni N, Pratesi G, Petrangolini G, Coradini D, et al. Isolation and *In vitro* Propagation of Tumorigenic Breast Cancer Cells with Stem/Progenitor Cell Properties. *Cancer Res [Internet]*. 2005 Jul 1 [cited 2018 Aug 30];65(13):5506-11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15994920>
407. He H, McHaney M, Hong J, Weiss ML. Cloning and Characterization of 3.1kb Promoter Region of the Oct4 Gene from the Fischer 344 Rat. *Open Stem Cell J [Internet]*. 2009 Jan 1 [cited 2018 Sep 2];1(1):30-9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22347989>
408. Welboren W-J, van Driel MA, Janssen-Megens EM, van Heeringen SJ, Sweep FC, Span PN, et al. ChIP-Seq of ER α and RNA polymerase II defines genes differentially responding to ligands. *EMBO J [Internet]*. 2009 May 20 [cited 2018 Sep 2];28(10):1418-28. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19339991>
409. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, et al. Epigenomic Analysis of

- Multilineage Differentiation of Human Embryonic Stem Cells. *Cell* [Internet]. 2013 May 23 [cited 2018 Sep 24];153(5):1134–48. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23664764>
410. Schmidt R, Plath K. The roles of the reprogramming factors Oct4, Sox2 and Klf4 in resetting the somatic cell epigenome during induced pluripotent stem cell generation. *Genome Biol* [Internet]. 2012 Oct 22 [cited 2018 Aug 25];13(10):251. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23088445>
 411. Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, et al. A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. *Cell* [Internet]. 2012 Dec 21 [cited 2018 Sep 24];151(7):1617–32. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23260147>
 412. Okamoto OK, Matheu A, Magnani L. Stem Cells in Translational Cancer Research. *Stem Cells Int* [Internet]. 2015 May 11 [cited 2018 Sep 24];2015:281072. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26078763>
 413. Magnani L, Stoeck A, Zhang X, Lanczky A, Mirabella AC, Wang T-L, et al. Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. *Proc Natl Acad Sci* [Internet]. 2013 Apr 16 [cited 2018 Sep 24];110(16):E1490–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23576735>
 414. Toh TB, Lim JJ, Chow EK-H. Epigenetics in cancer stem cells. *Mol Cancer* [Internet]. 2017 Dec 1 [cited 2018 Sep 24];16(1):29. Available from: <http://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-017-0596-9>
 415. Leis O, Eguiara a, Lopez-Arribillaga E, Alberdi MJ, Hernandez-Garcia S, Elorriaga K, et al. Sox2 expression in breast tumours and activation in breast cancer stem cells. *Oncogene*. 2012;31(11):1354–65.
 416. Jung K, Wang P, Gupta N, Gopal K, Wu F, Ye X, et al. Profiling gene promoter occupancy of Sox2 in two phenotypically distinct breast cancer cell subsets using chromatin immunoprecipitation and genome-wide promoter microarrays. *Breast Cancer Res*. 2014;16(6):1–13.
 417. Dontu G, Jackson KW, McNicholas E, Kawamura MJ, Abdallah WM, Wicha MS. Role of Notch signaling in cell-fate determination of human mammary stem/progenitor cells. *Breast Cancer Res* [Internet]. 2004 Dec 16 [cited 2018 Sep 24];6(6):R605. Available from:

- <http://www.ncbi.nlm.nih.gov/pubmed/15535842>
418. Cohen B, Shimizu M, Izrailit J, Ng NFL, Buchman Y, Pan JG, et al. Cyclin D1 is a direct target of JAG1-mediated Notch signaling in breast cancer. *Breast Cancer Res Treat* [Internet]. 2010 Aug 14 [cited 2018 Sep 24];123(1):113–24. Available from: <http://link.springer.com/10.1007/s10549-009-0621-9>
419. Meurette O, Stylianou S, Rock R, Collu GM, Gilmore AP, Brennan K. Notch activation induces Akt signaling via an autocrine loop to prevent apoptosis in breast epithelial cells. *Cancer Res* [Internet]. 2009 Jun 15 [cited 2018 Sep 24];69(12):5015–22. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19491273>
420. D'Angelo RC, Ouzounova M, Davis A, Choi D, Tchuenskam SM, Kim G, et al. Notch reporter activity in breast cancer cell lines identifies a subset of cells with stem cell activity. *Mol Cancer Ther* [Internet]. 2015 Mar 1 [cited 2018 Sep 24];14(3):779–87. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25673823>
421. Harrison H, Farnie G, Howell SJ, Rock RE, Stylianou S, Brennan KR, et al. Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. *Cancer Res* [Internet]. 2010 Jan 15 [cited 2018 Sep 24];70(2):709–18. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20068161>
422. Reedijk M, Odorcic S, Chang L, Zhang H, Miller N, McCready DR, et al. High-level Coexpression of JAG1 and NOTCH1 Is Observed in Human Breast Cancer and Is Associated with Poor Overall Survival. *Cancer Res* [Internet]. 2005 Sep 15 [cited 2018 Sep 24];65(18):8530–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16166334>
423. Baker A, Wyatt D, Bocchetta M, Li J, Filipovic A, Green A, et al. Notch-1-PTEN-ERK1/2 signaling axis promotes HER2+ breast cancer cell proliferation and stem cell survival. *Oncogene* [Internet]. 2018 Aug 10 [cited 2018 Sep 24];37(33):4489–504. Available from: <http://www.nature.com/articles/s41388-018-0251-y>
424. Lin X, Shang X, Manorek G, Howell SB. Regulation of the Epithelial-Mesenchymal Transition by Claudin-3 and Claudin-4. *PLoS One*. 2013;8(6).
425. Mendez MG, Kojima S-I, Goldman RD. Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition. *FASEB J*. 2010;24:1838–1851.
426. Trevor KT, McGuire JG, Leonova E V. Association of vimentin intermediate filaments with the centrosome. *J Cell Sci*. 1995;108 (Pt 1):343–56.

427. Manuel Iglesias J, Belouqui I, Garcia-Garcia F, Leis O, Vazquez-Martin A, Eguiara A, et al. Mammosphere Formation in Breast Carcinoma Cell Lines Depends upon Expression of E-cadherin. *PLoS One*. 2013;8(10):1–12.
428. Clark DW, Palle K. Aldehyde dehydrogenases in cancer stem cells: potential as therapeutic targets. *Ann Transl Med [Internet]*. 2016 Dec [cited 2018 Sep 2];4(24):518. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28149880>
429. Thiagarajan PS, Sinyuk M, Turaga SM, Mulkearns-Hubert EE, Hale JS, Rao V, et al. Cx26 drives self-renewal in triple-negative breast cancer via interaction with NANOG and focal adhesion kinase. *Nat Commun [Internet]*. 2018 Dec 8 [cited 2018 Sep 2];9(1):578. Available from: <http://www.nature.com/articles/s41467-018-02938-1>
430. Liu Y, Choi DS, Sheng J, Ensor JE, Liang DH, Rodriguez-Aguayo C, et al. HN1L Promotes Triple-Negative Breast Cancer Stem Cells through LEPR-STAT3 Pathway. *Stem Cell Reports [Internet]*. 2018 Jan 9 [cited 2018 Sep 2];10(1):212–27. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29249663>
431. Soady KJ, Kendrick H, Gao Q, Tutt A, Zvelebil M, Ordonez LD, et al. Mouse mammary stem cells express prognostic markers for triple-negative breast cancer. *Breast Cancer Res*. 2015;17(1):1–23.
432. Schwede M, Spentzos D, Bentink S, Hofmann O, Haibe-Kains B, Harrington D, et al. Stem Cell-Like Gene Expression in Ovarian Cancer Predicts Type II Subtype and Prognosis. *PLoS One*. 2013;8(3).
433. Merlos-Suárez A, Barriga FM, Jung P, Iglesias M, Céspedes MV, Rossell D, et al. The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell*. 2011;8(5):511–24.
434. Eppert K, Takenaka K, Lechman ER, Waldron L, Nilsson B, van Galen P, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med*. 2011 Sep;17(9):1086–93.
435. Lawson D a, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. 2015;
436. Huber-Keener KJ, Liu X, Wang Z, Wang Y, Freeman W, Wu S, et al. Differential Gene Expression in Tamoxifen-Resistant Breast Cancer Cells Revealed by a New Analytical Model of RNA-Seq Data. Culhane AC, editor. *PLoS One [Internet]*. 2012 Jul 23 [cited 2018 Aug 22];7(7):e41333. Available from: <http://dx.plos.org/10.1371/journal.pone.0041333>

437. Hurtado A, Holmes KA, Geistlinger TR, Hutcheson IR, Nicholson RI, Brown M, et al. Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen. *Nature* [Internet]. 2008 Dec 4 [cited 2018 Aug 22];456(7222):663–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19005469>
438. Bhatt S, Joshi S, Wu G, Katzenellenbogen BS. OCT-4: a novel estrogen receptor- α collaborator that promotes tamoxifen resistance in breast cancer cells. *Nat Publ Gr* [Internet]. 2016 [cited 2018 Aug 30];35:5722–34. Available from: www.nature.com/onc
439. Cho Y, Kang HG, Kim S-J, Lee S, Jee S, Ahn SG, et al. Post-translational modification of OCT4 in breast cancer tumorigenesis. *Cell Death Differ* [Internet]. 2018 Mar 6 [cited 2018 Sep 23];1. Available from: <http://www.nature.com/articles/s41418-018-0079-6>
440. Shaw AT, Kim D-W, Nakagawa K, Seto T, Crinó L, Ahn M-J, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med*. 2013;368:2385–94.
441. Michor F, Weaver VM. Understanding tissue context influences on intratumour heterogeneity. *Nat Cell Biol*. 2014 Apr;16(4):301–2.
442. Straussman R, Morikawa T, Shee K, Barzily-Rokni M, Qian ZR, Du J, et al. Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. *Nature*. 2012 Jul;487(7408):500–4.
443. Argent R, Kumari R, Clarke P, Onion D, Lobo D, Grabowska A, et al. Restoration of paracrine signalling within the tumour microenvironment increases tumour growth and activation of c-Met. In: 10th NCRI Cancer Conference: Poster Session C. Nottingham: NCRI; 2014. p. LB186.
444. Schmidt M, Böhm D, Von Törne C, Steiner E, Puhl A, Pilch H, et al. The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res*. 2008;68(13):5405–13.
445. Messersmith WA, Rajeshkumar N V, Tan AC, Wang XF, Diesl V, Choe SE, et al. Efficacy and pharmacodynamic effects of bosutinib (SKI-606), a Src/Abl inhibitor, in freshly generated human pancreas cancer xenografts. *Mol Cancer Ther* [Internet]. 2009 Jun 1 [cited 2018 Sep 4];8(6):1484–93. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19509264>
446. Marangoni E, Vincent-Salomon A, Auger N, Degeorges A, Assayag F, de Cremoux P, et al. A new model of patient tumor-derived breast cancer xenografts for preclinical assays. *Clin Cancer Res*. 2007 Jul;13(13):3989–98.
447. Gao H, Korn JM, Ferretti S, Monahan JE, Wang Y, Singh M, et al. High-throughput screening

- using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med*. 2015 Nov;21(11):1318–25.
448. Kudernatsch RF, Letsch A, Stachelscheid H, Volk HD, Scheibenbogen C. Doublets pretending to be CD34+ T cells despite doublet exclusion. *Cytom Part A* [Internet]. 2013 [cited 2018 Sep 13];83 A(2):173–6. Available from: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/cyto.a.22247>
 449. Kennedy A, Cribbs AP. Production and Concentration of Lentivirus for Transduction of Primary Human T Cells. In Humana Press, New York, NY; 2016 [cited 2018 Sep 13]. p. 85–93. Available from: http://link.springer.com/10.1007/978-1-4939-3753-0_7
 450. Zack JA, Kim SG, Vatakis DN. HIV restriction in quiescent CD4+ T cells. *Retrovirology* [Internet]. 2013 Apr 4 [cited 2018 Sep 13];10(1):37. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23557201>
 451. Buczacki SJA, Zecchini HI, Nicholson AM, Russell R, Vermeulen L, Kemp R, et al. Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature* [Internet]. 2013 Mar 27 [cited 2018 Aug 24];495(7439):65–9. Available from: <http://www.nature.com/articles/nature11965>
 452. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer Genome Landscapes. *Science* [Internet]. 2013 Mar 29;339(6127):1546–58. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3749880/>
 453. Neyman J. Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability. [Internet]. University of California Press; 1967 [cited 2018 Sep 12]. Available from: <https://projecteuclid.org/euclid.bsmsp/1200512992>
 454. Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognit Lett* [Internet]. 2010 Jun 1 [cited 2018 Sep 13];31(8):651–66. Available from: <https://www.sciencedirect.com/science/article/pii/S0167865509002323>
 455. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Publ Gr* [Internet]. 2016 [cited 2018 Sep 12];13. Available from: www.nature.com/nrclinonc
 456. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* [Internet]. 2012 Jun 4 [cited 2018 Sep 12];486(7403):395–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22495314>

457. DuPage M, Dooley AL, Jacks T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat Protoc* [Internet]. 2009 Jun 25;4(7):1064–72. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2757265/>
458. Vara JA, Portela A, Ortín J, Jiménez A. Expression in mammalian cells of a gene from *Streptomyces alboniger* conferring puromycin resistance. *Nucleic Acids Res* [Internet]. 1986 Jun 11 [cited 2018 Sep 14];14(11):4617–24. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/3714487>
459. Kita-Matsuo H, Barcova M, Prigozhina N, Salomonis N, Wei K, Jacot JG, et al. Lentiviral Vectors and Protocols for Creation of Stable hESC Lines for Fluorescent Tracking and Drug Resistance Selection of Cardiomyocytes. Blagosklonny M V., editor. *PLoS One* [Internet]. 2009 Apr 8 [cited 2018 Sep 14];4(4):e5046. Available from: <http://dx.plos.org/10.1371/journal.pone.0005046>
460. Levitin HM, Yuan J, Sims PA. Single-Cell Transcriptomic Analysis of Tumor Heterogeneity. *Trends in cancer* [Internet]. 2018 Apr 1 [cited 2018 Sep 26];4(4):264–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29606308>
461. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, et al. The evolutionary history of lethal metastatic prostate cancer. *Nature* [Internet]. 2015 Apr 1 [cited 2018 Sep 26];520(7547):353–7. Available from: <http://www.nature.com/articles/nature14347>
462. Alderton GK. Spreading the seed. *Nat Rev | CANCER* [Internet]. 2015 [cited 2018 Sep 26];15. Available from: <http://dx.doi.org/10.1038/>
463. Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015. *JAMA Oncol* [Internet]. 2017 Apr 1 [cited 2018 Sep 23];3(4):524. Available from: <http://oncology.jamanetwork.com/article.aspx?doi=10.1001/jamaoncol.2016.5688>
464. Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* [Internet]. 2016 Oct 8 [cited 2018 Sep 23];388(10053):1459–544. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27733281>
465. King A, Broggio J. Statistical bulletin: Cancer registration statistics, England: 2016

- [Internet]. Office for National Statistics. 2018 [cited 2018 Sep 24]. Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/final2016>
466. Pepper JW, Scott Findlay C, Kassen R, Spencer SL, Maley CC. SYNTHESIS: Cancer research meets evolutionary biology. *Evol Appl* [Internet]. 2009 Jan 27 [cited 2018 Sep 24];2(1):62–70. Available from: <http://doi.wiley.com/10.1111/j.1752-4571.2008.00063.x>
 467. Merlo LMF, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. *Nat Rev Cancer* [Internet]. 2006 Dec 16 [cited 2018 Sep 24];6(12):924–35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17109012>
 468. Aktipis CA, Nesse RM. Evolutionary foundations for cancer biology. *Evol Appl* [Internet]. 2013 Jan 1 [cited 2018 Sep 24];6(1):144–59. Available from: <http://doi.wiley.com/10.1111/eva.12034>
 469. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature* [Internet]. 2009 Apr 9 [cited 2018 Sep 24];458(7239):719–24. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19360079>
 470. Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet* [Internet]. 2012 Nov [cited 2018 Sep 24];13(11):795–806. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23044827>
 471. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* [Internet]. 2012 Jun 18 [cited 2018 Sep 23];486(7403):346–52. Available from: <http://www.nature.com/articles/nature10983>
 472. Perez EA. Breast Cancer Management: Opportunities and Barriers to an Individualized Approach. *Oncologist* [Internet]. 2011 Jan 1 [cited 2018 Sep 24];16(Supplement 1):20–2. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21278437>
 473. Cai S, Geng S, Jin F, Liu J, Qu C, Chen B. POU5F1/Oct-4 expression in breast cancer tissue is significantly associated with non-sentinel lymph node metastasis. 2010 [cited 2018 Aug 30]; Available from: <http://nomograms.mskcc.org/Breast/Breast>
 474. Cruz C, Castroviejo-Bermejo M, Gutiérrez-Enríquez S, Llop-Guevara A, Ibrahim YH, Gris-Oliver A, et al. RAD51 foci as a functional biomarker of homologous recombination repair and PARP inhibitor resistance in germline BRCA-mutated breast cancer. *Ann Oncol*

- [Internet]. 2018 May 1 [cited 2018 Sep 24];29(5):1203–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29635390>
475. Raina R, Battle A, Lee H, Packer B, Ng AY. Self-taught learning. In: Proceedings of the 24th international conference on Machine learning - ICML '07 [Internet]. New York, New York, USA: ACM Press; 2007 [cited 2018 Sep 27]. p. 759–66. Available from: <http://portal.acm.org/citation.cfm?doid=1273496.1273592>
 476. Haricharan S, Lei J, Ellis M. Mammary Ductal Environment Is Necessary for Faithful Maintenance of Estrogen Signaling in ER⁺ Breast Cancer. *Cancer Cell* [Internet]. 2016 Mar 14 [cited 2019 Jan 28];29(3):249–50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26977876>
 477. Sflomos G, Dormoy V, Metsalu T, Jeitziner R, Battista L, Scabia V, et al. A Preclinical Model for ER⁺-Positive Breast Cancer Points to the Epithelial Microenvironment as Determinant of Luminal Phenotype and Hormone Response. *Cancer Cell*. 2016 Mar;29(3):407–22.
 478. Ghosh A, Sarkar S, Banerjee S, Behbod F, Tawfik O, McGregor D, et al. MIND model for triple-negative breast cancer in syngeneic mice for quick and sequential progression analysis of lung metastasis. [cited 2019 Jan 28]; Available from: <https://doi.org/10.1371/journal.pone.0198143>
 479. Mohibi S, Mirza S, Band H, Band V. Mouse models of estrogen receptor-positive breast cancer. *J Carcinog* [Internet]. 2011 [cited 2019 Jan 28];10:35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22279420>
 480. Medina D, Butel JS, Socher SH, Miller FL. Mammary tumorigenesis in 7,12-dimethylbenzanthracene-treated C57BL x DBA/2f F1 mice. *Cancer Res* [Internet]. 1980 Feb 1 [cited 2019 Jan 28];40(2):368–73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/6243251>
 481. MEDINA D, KITTRELL FS, SHEPARD A, STEPHENS LC, JIANG C, LU J, et al. Biological and genetic properties of the p53 null preneoplastic mammary epithelium. *FASEB J* [Internet]. 2002 Jun 10 [cited 2019 Jan 28];16(8):881–3. Available from: <http://www.fasebj.org/doi/10.1096/fj.01-0885fje>
 482. Chan SR, Vermi W, Luo J, Lucini L, Rickert C, Fowler AM, et al. STAT1-deficient mice spontaneously develop estrogen receptor α -positive luminal mammary carcinomas. *Breast Cancer Res* [Internet]. 2012 Feb 20 [cited 2019 Jan 28];14(1):R16. Available from:

- <http://www.ncbi.nlm.nih.gov/pubmed/22264274>
483. Hinohara K, Wu H-J, Vigneau S, McDonald TO, Igarashi KJ, Yamamoto KN, et al. KDM5 Histone Demethylase Activity Links Cellular Transcriptomic Heterogeneity to Therapeutic Resistance. *Cancer Cell* [Internet]. 2018 Dec 10 [cited 2019 Jan 28];34(6):939–953.e9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30472020>
484. Echeverria GV, Seth S, Ge Z, Carugo A, Bristow C, Mundi P, et al. Abstract 212: High-resolution barcoding in patient-derived xenografts of triple-negative breast cancer reveals reversible chemoresistance conferred by non-mutational mechanisms. *Cancer Res* [Internet]. 2018 Jul 1 [cited 2019 Jan 28];78(13 Supplement):212–212. Available from: <http://cancerres.aacrjournals.org/lookup/doi/10.1158/1538-7445.AM2018-212>